ABSTRACT

Title of Dissertation: SPURIOUS LYAPUNOV EXPONENTS COMPUTED USING THE ECKMANN-RUELLE PROCEDURE

Joshua A. Tempkin, Doctor of Philosophy, 1999

Dissertation directed by: Professor James A. Yorke
Department of Mathematics

**Abstract**

Lyapunov exponents, important invariants of a complicated dynamical process, can be difficult to determine from experimental data. In particular, when using embedding theory to build chaotic attractors in a reconstruction space, extra "spurious" Lyapunov exponents can arise that are not Lyapunov exponents of the original system. By studying the local linearization matrices that are key to a popular method for computing Lyapunov exponents, we determine explicit formulas for the spurious exponents in certain cases. Notably, when a two-dimensional system with Lyapunov exponents $\alpha$ and $\beta$ is reconstructed in a five-dimensional space, we show that the reconstructed system has exponents $\alpha$, $\beta$, $2\alpha$, $2\beta$, and $\alpha + \beta$.

# SPURIOUS LYAPUNOV EXPONENTS COMPUTED USING THE ECKMANN-RUELLE PROCEDURE

by

Joshua A. Tempkin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland at College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1999

Advisory Committee:

Professor James A. Yorke, Chairman/Advisor
Professor Celso Grebogi
Professor Edward Ott
Professor J. Robert Dorfman
Professor Denny Gulick

# DEDICATION

To Margaret, for waiting,
and
To Andrea, for not waiting

# ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. James Yorke, for his support and guidance throughout this thesis project. He has taught me a lot about mathematics and how to communicate ideas. His insistence that proofs be unencumbered by notation helped me numerous times to clarify my ideas. His intuition is astonishing and his nurturing patience seems limitless. I have grown tremendously since we began working together.

I also would like to thank my thesis committee, Dr. Celso Grebogi, Dr. Edward Ott, Dr. Robert Dorfman, and Dr. Denny Gulick, for being available and helpful whenever I had questions. In addition, I had several interesting discussions with Dr. Brian Hunt and with Dr. Tim Sauer of George Mason University. Dr. Sauer was especially helpful with the work in Appendix B.

During my first few years of graduate study, Dr. Robert Warner and Dr. Frances Gulick took me under their wings and encouraged me. Millie Stengel and the secretaries in the Math Department have helped in many many ways, big and small, over the years.

The students and postdocs in the Chaos Group have been very supportive. They are a great bunch of people, one and all, and several deserve special mention: Linda Moniz, Dj Patil, Myong-Hee Sung, Mike Roberts, Mitrajit Dutta, and Guocheng Yuan.

I also want thank to my parents. Their unwavering faith in me and their support, encouragement, and love have made a big difference in my life.

Finally, special thanks, heartfelt gratitude, and warm appreciation go to my dear and loving wife, Margaret, for *absolutely everything*. She has put up with more from me than I will admit to, and still she comes away smiling. No more late nights in the basement, honey, I promise! Now I can get a "real" job!

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

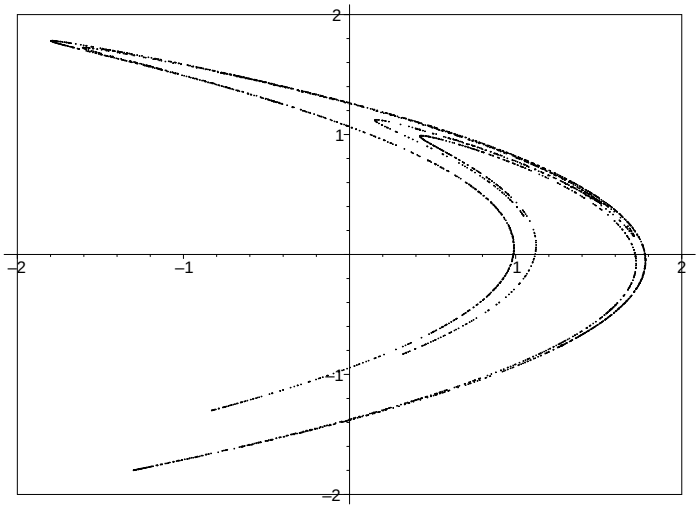
# Chapter 1

# Introduction

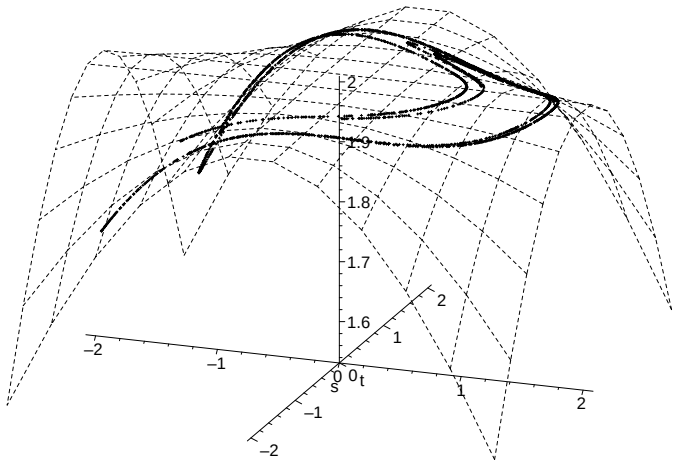
In the analysis of observed physical systems, it has become standard practice to study an auxiliary system reconstructed from a time series of measured data. Successful reconstruction of the original system's attractor is the basis of the method for Lyapunov exponent calculation proposed by Eckmann and Ruelle [1, 2] and by Sano and Sawada [3]. However, since the reconstructed attractor often lies in a larger dimensional space than the original system, the calculations produce too many exponents. This leads to an important question: how do we distinguish the true Lyapunov exponents of the underlying system from the extra "spurious" ones present for the reconstructed system? We answer this question for two specific cases: (a) when a one-dimensional system is reconstructed in $m$ dimensions and (b) when a two-dimensional system is reconstructed in five dimensions.

The attractor reconstruction process begins by choosing a number $m$ and observing the present state $p$ of the underlying system with that number $m$ of independent measurements $\pi_i(p)$, $i = 1, \ldots, m$. For each point $p$ in the phase space, there is an $m$-dimensional vector $\pi(p) = (\pi_1(p), \ldots, \pi_m(p))$. This produces a **measurement function** $\pi$ that associates points in $\mathbb{R}^m$ with points in the phase space of the underlying dynamical system. In practice, the measurement function $\pi$ often consists of time-delayed versions of a scalar measurement (see Takens and others [4, 5]). Under certain genericity conditions (see, for example, [6]), the original dynamical phase space attractor $A$ will be topologically equivalent to its reconstructed image $\pi(A)$ in $m$-dimensional Euclidean space. See Figure 1.1.

The set of measurement vectors $\pi(p)$ in the reconstructed attractor $\pi(A)$ can be studied for geometrical and dynamical properties. Looking to the reconstructed attractor for dynamical properties of the original attractor was suggested in 1985 by Eckmann and Ruelle, et. al. [1, 2] and also by Sano and Sawada [3]. It is usually necessary that $m$ be chosen large enough that there is a one-to-one correspondence between points of the original attractor and points of the reconstructed attractor [6]. This requirement often forces the reconstruction space




(a)




(b)

Figure 1.1: The measurement function $\pi$ reconstructs the true attractor $A$ of the underlying dynamics (shown in (a)) as the set $\pi(A)$ in some $\mathbb{R}^m$ (shown in (b)). When the reconstruction dimension $m$ is large enough, the generic (smooth) measurement function $\pi$ will be a one-to-one correspondence between $A$ and $\pi(A)$.

to have larger dimension than the original system. In these cases, Lyapunov exponent calculations in the reconstruction space produce $m$ real numbers, not all of which can be Lyapunov exponents of the original dynamical system. For example, if a two-dimensional dynamical system is reconstructed in five dimensions, current methods compute five "exponents" in the reconstruction space. At most two of these can be Lyapunov exponents of the original system; the other numbers are "spurious" exponents.

How, then, do we tell the true exponents from the spurious exponents? Parlitz [7] proposed that recomputing the exponents using the reversed time series would make the true exponents switch sign. Bryant, Brown, and Abarbanel looked at the local "thickness" of the data set to identify spurious exponents [8]. In addition, some authors have proposed removing the extra dimensions by projecting the reconstructed dynamics to its tangent plane (see [9, pp. 336–339] and [10, pp. 2156-2157]. In the present paper, we study the original Eckmann-Ruelle algorithm in order to clarify its output.

The algorithm presented in [2] for computing Lyapunov exponents has three major steps. First, one reconstructs the attractor in some $m$-dimensional Euclidean space of measurements as indicated above. On the reconstructed attractor, there is a time-$\tau$ map $F$ which takes the $m$-vector $P_t$ at time $t$ to the $m$-vector $P_{t+\tau}$ at time $t+\tau$. This map $F$ represents the reconstructed dynamics, and $F(P_t)$ is defined to be $P_{t+\tau}$. In the second step, one computes a local linearization matrix for $F$ at each point $P$ of the reconstructed attractor by finding the $m \times m$ matrix $M$ (depending on $P$) which satisfies as closely as possible

$$P_{t+\tau} - F(P) \approx M(P_t - P) \qquad \text{for all } t \text{ for which } P_t \text{ is close to } P.$$

We call $M = M(P)$ the Eckmann-Ruelle linearization at $P$. In the last step (which will be discussed in more detail in Chapter 3), one computes the Lyapunov exponents of $F$ from a matrix product of these local linearizations. That is, the Lyapunov exponents will be the various values achieved by

$$\lim_{n\to\infty} \frac{1}{n} \ln \|M_{n-1} M_{n-2} \cdots M_1 M_0 \nu\|$$

for various unit vectors $\nu \in \mathbb{R}^m$, where $M_i = M(P_i)$ is the best local linearization matrix (i.e., the Eckmann-Ruelle linearization) at the point $P_i = \pi(p_i)$ in the trajectory.

To understand the output from this algorithm, it is crucial to determine the Eckmann-Ruelle linearizations. At first glance, one might think these linearizations should produce derivative matrices, $DF_P$. This, however, will not happen in general. Suppose that an attractor reconstructed in $m$-dimensional Euclidean space lies within a lower-dimensional surface in $\mathbb{R}^m$. The reconstructed dynamics are well-defined on this surface, but they are not defined off of it, and so the

classical $m \times m$ derivative matrix $DF_P$ will not exist. See Figure 1.1b. Therefore, the local linearizations cannot be derivative matrices.

In this paper, we study the local linearization matrices. For the two cases mentioned above, we will show that these matrices have several important properties. First and foremost, the linearizations of the reconstructed dynamics are surprisingly good in the following sense (see Theorem 2.2 in Chapter 2). For any matrix $L$, if we "linearize" the reconstructed dynamics $F$ about the point $P$, we obtain for some integer $k$

$$F(P_t) - F(P) - L\,(P_t - P) = O\Big(\|P_t - P\|^k\Big) \quad \text{as } P_t \to P. \tag{1.1}$$

We write $g(x) = O(|x|^k)$ as $x \to 0$ to mean that there exists some constant $C$ such that $\|g(x)\| \leq C|x|^k$ for all $x$ in some neighborhood of 0. For most choices of $L$, we expect $k = 1$ in (1.1). For the traditional derivative $L = DF_P$, we expect $k = 2$ (when it exists). However, the Eckmann-Ruelle linearization $L = M(P)$ in our specific cases has $k$ larger than this. In the case of a one-dimensional system reconstructed in $m$ dimensions, we find $k = m + 1$, and in the case of a two-dimensional system reconstructed in 5 dimensions, we find $k = 3$. The second important property of these matrices is that there are natural coordinate systems with respect to which the linearization matrices have a special upper-triangular matrix representation. This upper triangular form allows us to easily compute the Lyapunov exponents (which depend on the diagonal entries when the matrices are upper triangular). Moreover, for generic measurement functions, these exponents will be completely independent of the specific measurement function $\pi$ used in the reconstruction, depending only on the dynamics of the original system. This key property allows us to derive explicit formulas for the Lyapunov exponents produced by the Eckmann-Ruelle procedure in the low noise limit. We will demonstrate this in Chapter 3.

Our paper is structured in the following manner. In Chapter 2, we determine the linear map that provides the best local linearization to the reconstructed dynamics. In Chapter 3, we give the matrix representation alluded to above and derive formulas for the Lyapunov exponents produced by the algorithm. The goal of Chapter 4 is to prove that numerical determinations of the local linearization matrix converge to the Eckmann-Ruelle matrix as the radius of the neighborhood shrinks to zero. Chapter 5 presents numerical work illustrating our theoretical results. Finally, appendices are included which contain technical lemmas used in the text.

# Chapter 2

# Local Linearizations

A simple example where $f(p) = 2p(mod\ 2\pi)$ on the interval $[0, 2\pi]$ allows us to describe the general problem. We reconstruct this interval in $\mathbb{R}^2$ via the measurement function $\pi(p) = (\cos(p), \sin(p))$. The observed dynamics $F = \pi f \pi^{-1}$ maps points on the unit circle: $\pi(p)$ to $\pi(f(p))$. The following question is central to our discussion: for any given point $P = \pi(p)$ on the unit circle $\pi([0, 2\pi])$, which $2 \times 2$ matrix provides the best local linearization of $F$ around $P$ (in the sense of (1.1))? Since $F$ is not defined off the unit circle, the traditional derivative of $F$ does not exist. On the other hand, the linearizations of $F$ are $2 \times 2$ matrices while the derivatives of the original system $f$ are $1 \times 1$. It follows that no linearization of $F$ can be a derivative matrix. What, then, is the best local linearization of $F$ near $P$ (if one even exists)?

In this chapter, we elucidate the nature of the local linearization matrices for the reconstructed dynamics. In their papers [1, 2], Eckmann and Ruelle discussed the best local linearization of the reconstructed dynamics. We will define the "Eckmann-Ruelle linearization" to be the unique linear map with certain properties. In Theorems 2.2 and 2.6, we will show that our definition provides the best local linearization.

In the simple example above, we specified the measurement function $\pi$ explicitly. In practice, however, one rarely knows the measurement function that arises from the attractor reconstruction process. Time-delay embeddings are commonly used, but even then, one cannot know the measurement function completely without knowledge of how the scalars in the time-series relate to the states of the system. This knowledge is often unavailable in experimental situations.

For this reason, we focus on measurement functions with generic properties (generic in the sense of prevalence). A property is **generic** in the sense of prevalence if whenever a function lacks the property, arbitrarily small perturbations of that function have the property with probability 1 [11]. In the situations that interest us, the generic measurement function can be taken to be a smooth diffeomorphism from the underlying phase space into $m$-dimensional Euclidean space $\mathbb{R}^m$. The genericity conditions that guarantee the (topological) equivalence of

the underlying and reconstructed phase space attractors [6] typically force the reconstruction space to have larger dimension than the underlying attractor. In fact, the dimension $m$ of the reconstruction space can be at least twice that of the underlying system.

We adopt the convention that lower-case letters refer to the underlying system while upper-case letters refer to the reconstructed system. For example, $P = \pi(p)$ is the point in the reconstructed phase space $\mathbb{R}^m$ corresponding to the point $p$ in the underlying phase space. The convention will also extend to sets in the various spaces: $u$ could be a neighborhood of $p$ in the underlying phase space while $U$ could be a neighborhood of $P$ in the reconstructed phase space. We represent the dynamical flow on the underlying phase space by the time-$\tau$ map $f$ for some $\tau$. The measurement function $\pi$ maps the underlying phase space into $\mathbb{R}^m$ for some $m$. In $\mathbb{R}^m$, there is an induced time-$\tau$ flow map $F = \pi f \pi^{-1}$ that maps each $\pi(p)$ to $\pi(f(p))$, and we refer to $F$ as the reconstructed dynamics. Assuming that $\pi$ is a one-to-one correspondence on the underlying phase space (i.e., that $p \neq q$ implies $\pi(p) \neq \pi(q)$), the map $F$ is well-defined on the reconstructed phase space.

We examine the special case when the underlying map $f$ is one-dimensional and the following specific assumptions about $f$, $\pi$, and a point $P = \pi(p)$ hold.

**(A1)** *The underlying dynamical system $f$ maps the unit interval $[0,1]$ into itself. In addition, we assume that $f$ has $m+1$ continuous derivatives, i.e., $f$ is $C^{m+1}$.*

**(A2)** *The measurement function $\pi$ maps $[0,1]$ into $\mathbb{R}^m$, and $\pi$ is $C^{m+1}$.*

**(A3)** *For this $p \in [0,1]$, $P = \pi(p) \in \mathbb{R}^m$, and the set $\pi^{-1}(P)$ contains only one point, namely $p$.*

**(A4)** *The first $m$ derivative vectors for $\pi$ at $p$, i.e., $\pi^{(n)}(p) := \frac{d^n \pi}{dp^n}(p)$ for $n = 1, 2, \ldots, m$, are linearly independent in $\mathbb{R}^m$.*

Properties (A3) and (A4) are generic properties in the space of $C^{m+1}$ functions from $[0,1]$ into $\mathbb{R}^m$. By (A4), the derivative vector $\pi'(p)$ is nonzero. Together with (A3) and the Inverse Function Theorem, this implies

**(A3')** *There are neighborhoods $u \subseteq [0,1]$ of $p$ and $U \subseteq \pi([0,1]) \subset \mathbb{R}^m$ of $P$ such that $\pi(x) \in U$ if and only if $x \in u$, and $\pi|_u$ is a diffeomorphism.*

This statement has a useful consequence. Since $\pi|_u$ is a diffeomorphism, there is a constant $C_\pi > 1$ such that if $P_1, P_2 \in U$ with $P_i = \pi(p_i)$, $p_i \in u$, $i = 1, 2$, then

$$\frac{1}{C_\pi} \|P_1 - P_2\| \leq |p_1 - p_2| \leq C_\pi \|P_1 - P_2\|. \tag{2.1}$$

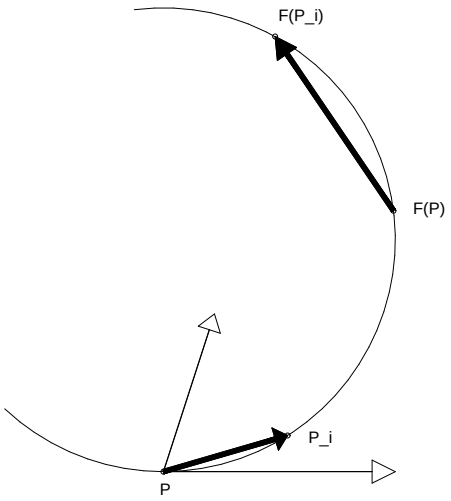



Figure 2.1: The derivative vectors $\pi^{(1)}(p)$ and $\pi^{(2)}(p)$ (thin arrows) form the canonical embedding basis for $\mathbb{R}^2$ at $P = \pi(p)$. Also, the small tangent vector $\Delta P_i = P_i - P$ maps to its image vector $F(P_i) - F(P)$ near $F(P)$ (thick arrows).

This inequality will be useful in translating statements back and forth between the underlying space and the reconstruction space.

We begin our analysis of this case by examining the Taylor expansion of $\pi$ about a point $P = \pi(p)$ in $m$-space for which (A3) and (A4) hold. Since $\pi$ is one-to-one on the neighborhood $u$, any point $P + \Delta P$ in $\pi(u)$ near $P$ can be written in the form

$$P + \Delta P = \pi(p + h) = P + \sum_{n=1}^{m} \frac{\pi^{(n)}(p)}{n!} h^n + Rem(h)$$

where $Rem(h)$ is the Taylor remainder term (an $m$-vector here). There is a constant $C_{Taylor} > 0$ such that

$$\|Rem(h)\| \leq C_{Taylor} |h|^{m+1}.$$

By hypothesis (A4), the vectors $\pi^{(1)}(p), \ldots, \pi^{(m)}(p)$ are linearly independent and form a basis for $\mathbb{R}^m$ which we call the **canonical embedding basis** at $P$. (See Figure 2.1.) The little vector $\Delta P$ at $P$ can be written conveniently in this basis:

$$\Delta P = \left( h, \frac{1}{2} h^2, \ldots, \frac{1}{m!} h^m \right)_P + Rem(h). \tag{2.2}$$

We look at the image $F(P) = \pi(f(p))$ in essentially the same way. The Taylor expansion of $F(P + \Delta P) = F(\pi(p + h)) = \pi(f(p + h))$ is given by

$$F(P + \Delta P) = F(P) + \sum_{n=1}^{m} \frac{(\pi \circ f)^{(n)}(p)}{n!} h^n + Rem_f(h).$$

Here, $Rem_f$ is the Taylor remainder vector associated with $\pi \circ f$. Without loss of generality, we can choose the constant $C_{Taylor}$ so that we also have

$$\|Rem_f(h)\| \leq C_{Taylor}|h|^{m+1}.$$

At this point, we are ready to define the Eckmann-Ruelle linearization. The definition will be justified by the properties stated in Theorem 2.2.

**Definition 2.1.** *Assuming (A1) – (A4), we define the* **Eckmann-Ruelle linearization** $M = M(P)$ *to be the unique linear map from* $\mathbb{R}^m$ *to* $\mathbb{R}^m$ *satisfying*

$$M\ \pi^{(n)}(p) = (\pi \circ f)^{(n)}(p) \qquad \text{for each } n = 1, \ldots, m. \tag{2.3}$$

$M$ is well-defined and unique because the set of vectors $\{\pi^{(1)}(p), \ldots, \pi^{(m)}(p)\}$ on the left-hand side of (2.3) forms a basis for $\mathbb{R}^m$ by assumption (A4). While it may be convenient to think of $M$ as a matrix, no coordinate system has yet been specified. We will give a matrix representation for $M$ in the next chapter. Now, we prove that $M(P)$ is in fact the best linear approximation to the reconstructed dynamics $F$ near $P$.

We say that $g(x) = O(|x|^k)$ as $x \to 0$ if there exists a constant $C$ such that $\|g(x)\| \leq C|x|^k$ for all $x$ in some neighborhood of 0. Note that the error term in (2.4) of Theorem 2.2 can be far smaller than that for the usual Jacobian matrix, which would be $O(\|\Delta P\|^2)$.

**Theorem 2.2 (Local Linearization, $\mathbb{R}^1 \to \mathbb{R}^m$).** *Assume (A1) and (A2). Let* $P = \pi(p)$ *be a point of* $\pi([0,1])$ *for which (A3) and (A4) hold. If* $P$ *is in the closure of a trajectory of* $F$, $P_0, P_1, \cdots \in \mathbb{R}^m$, *then the Eckmann-Ruelle linearization* $M = M(P)$ *defined by (2.3) is the unique linear map such that*

$$\begin{aligned} F(P+\Delta P) - F(P) - M\Delta P &= O\Big(\|\Delta P\|^{m+1}\Big) \\ \text{as } \|\Delta P\| \to 0, &\qquad \text{where } P + \Delta P \in \pi([0,1]). \end{aligned} \tag{2.4}$$

**Proof.** For small $\Delta P$ with $P + \Delta P \in \pi(u)$, we can write $P + \Delta P = \pi(p + h)$. Note that $F(P) = \pi(f(p))$ and $F(P + \Delta P) = \pi(f(p + h))$. First, we prove that the Eckmann-Ruelle linearization $M = M(P)$ satisfies (2.4). Using the definition of $M$ in equation (2.3), we cancel terms from the two Taylor series:

$$\begin{aligned} &F(P+\Delta P) - F(P) - M\Delta P \\ &\quad = \pi(f(p+h)) - \pi(f(p)) - M\left(\pi(p+h) - \pi(p)\right) \\ &\quad = \left(\sum_{n=1}^{m} (\pi \circ f)^{(n)}(p)\frac{h^n}{n!} + Rem_f(h)\right) - M\left(\sum_{n=1}^{m} \pi^{(n)}(p)\frac{h^n}{n!} + Rem(h)\right) \\ &\quad = Rem_f(h) - MRem(h). \end{aligned}$$

Since $M$ is a fixed map and $|h| \leq C_\pi \|\Delta P\|$ by (2.1), we have

$$\begin{aligned} \|F(P+\Delta P) - F(P) - M\Delta P\| &\leq \|Rem_f(h)\| + \|MRem(h)\| \\ &\leq C_{Taylor}(1+\|M\|)|h|^{m+1} \\ &\leq C_{Taylor}(1+\|M\|)C_\pi^{m+1}\|\Delta P\|^{m+1} \end{aligned}$$

for all $\|\Delta P\|$ sufficiently small, establishing (2.4).

It remains to show that $M$ is the only linear map that satisfies (2.4). Suppose to the contrary there is another linear map $\widetilde{M}$ such that

$$\left\|F(P+\Delta P) - F(P) - \widetilde{M}\Delta P\right\| \leq C_1\|\Delta P\|^{m+1} \qquad \text{as } \Delta P \to 0,\ P+\Delta P \in \pi(u)$$

for some constant $C_1$. Set $A := \widetilde{M} - M$. For small enough $\Delta P$ with $P+\Delta P \in \pi(u)$,

$$\begin{aligned} \|A\Delta P\| &\leq \|F(P+\Delta P) - F(P) - M\Delta P\| + \left\|F(P+\Delta P) - F(P) - \widetilde{M}\Delta P\right\| \\ &\leq C_2\|\Delta P\|^{m+1} \end{aligned}$$

where $C_2 := C_1 + C_{Taylor}(1+\|M\|)C_\pi^{m+1}$. With respect to the canonical embedding basis at $P$, the vector $\Delta P$ has the form in (2.2), and so,

$$\begin{aligned} \left\|A\left(h, \frac{1}{2}h^2, \ldots, \frac{1}{m!}h^m\right)_P\right\| &\leq \|A\Delta P\| + \|ARem(h)\| \\ &\leq C_2\|\Delta P\|^{m+1} + \|A\|\ C_{Taylor}|h|^{m+1} \\ &\leq C_2 C_\pi^{m+1}|h|^{m+1} + \|A\|\ C_{Taylor}|h|^{m+1} \\ &:= C_3|h|^{m+1} \end{aligned}$$

Therefore, for $h \in u$ sufficiently small

$$\left\|A\left(h, \frac{1}{2}h^2, \ldots, \frac{1}{m!}h^m\right)_P\right\|^2 \leq C_3^2|h|^{2m+2}.$$

If we represent the matrix $A$ with respect to the canonical embedding basis at $P$, then the left-hand side is a polynomial in $h$, call it $p(h)$, with degree at most $2m$. Since $P$ is a limit point of the trajectory in $\mathbb{R}^m$, there are infinitely many $h_i \in u$, $P + \Delta P_i = \pi(p + h_i) \in U$, $h_i \to 0$ for which this polynomial satisfies $|p(h_i)| \leq C_3^2|h_i|^{2m+2}$. By Proposition A.1 of Appendix A, $p(h)$ must be the zero polynomial. This, in turn, forces all the elements of the matrix $A$ to be zero. Thus, $A = 0$ and $\widetilde{M} = M$, proving that $M = M(P)$ is indeed the unique linear map satisfying (2.4). □

To make this method explicit (and for later use), we compute the Eckmann-Ruelle linearization for our doubling map example. Let $f(p) = 2p(mod\ 2\pi)$ on

$[0, 2\pi]$ and $\pi(p) = (\cos(p), \sin(p))$. Let $0 \le p_0 < 2\pi$ and set $\pi(p_0) = (x_0, y_0)$ where $x_0 = \cos(p_0)$ and $y_0 = \sin(p_0)$. Then,

$$\pi^{(1)}(p_0) = \begin{pmatrix} -y_0 \\ x_0 \end{pmatrix} \quad \text{and} \quad \pi^{(2)}(p_0) = \begin{pmatrix} -x_0 \\ -y_0 \end{pmatrix}.$$

Also, we have $(\pi \circ f)(p) = (\cos(2p), \sin(2p))$ and

$$(\pi \circ f)^{(1)}(p_0) = \begin{pmatrix} -2\,(2x_0y_0) \\ 2\,(x_0^2 - y_0^2) \end{pmatrix} \quad \text{and} \quad (\pi \circ f)^{(2)}(p_0) = \begin{pmatrix} -4\,(x_0^2 - y_0^2) \\ -4\,(2x_0y_0) \end{pmatrix}.$$

Since $M\pi^{(1)}(p_0) = (\pi \circ f)^{(1)}(p_0)$ and $M\pi^{(2)}(p_0) = (\pi \circ f)^{(2)}(p_0)$ by (2.3), we are led to the matrix equation

$$M \begin{pmatrix} -y_0 & -x_0 \\ x_0 & -y_0 \end{pmatrix} = \begin{pmatrix} -2\,(2x_0y_0) & -4\,(x_0^2 - y_0^2) \\ 2\,(x_0^2 - y_0^2) & -4\,(2x_0y_0) \end{pmatrix}$$

which we can solve for $M$:

$$M(x_0, y_0) = \begin{pmatrix} 4x_0^3 & -4y_0^3 \\ 2y_0\,(2x_0^2 + 1) & 2x_0\,(2y_0^2 + 1) \end{pmatrix}.$$

In Chapter 5, we will describe numerical experiments for this example showing that the linearization matrix calculated by the computer is close to this $M(x_0, y_0)$.

We next examine the special case when a two-dimensional underlying system $f$ is reconstructed into five-dimensional space. We make the following specific assumptions about $f$, $\pi$, and a point $P = \pi(p)$ in $\mathbb{R}^5$.

**(B1)** *$f$ maps the unit square $[0, 1] \times [0, 1]$ into itself, and $f$ is $C^3$.*

**(B2)** *$\pi$ maps $[0, 1] \times [0, 1]$ into $\mathbb{R}^5$, and $\pi$ is $C^3$.*

**(B3)** *For this $p \in [0, 1] \times [0, 1]$, $P = \pi(p) \in \mathbb{R}^5$, and the set $\pi^{-1}(P)$ contains only one point, namely $p$.*

**(B4)** *The first and second order partial derivative vectors for $\pi$ at $p$, namely*

$$\pi_x(p) := \frac{\partial \pi}{\partial x}(p), \qquad \pi_y(p) := \frac{\partial \pi}{\partial y}(p)$$

$$\pi_{xx}(p) := \frac{\partial^2 \pi}{\partial x^2}(p), \qquad \pi_{xy}(p) := \frac{\partial^2 \pi}{\partial x \partial y}(p), \qquad \pi_{yy}(p) := \frac{\partial^2 \pi}{\partial^2 y}(p)$$

*are linearly independent in $\mathbb{R}^5$.*

Properties (B3) and (B4) are generic properties in the space of $C^3$ functions from $[0, 1] \times [0, 1]$ into $\mathbb{R}^5$. As in the previous case, we obtain a seemingly stronger statement as a consequence of (B3), (B4), and the Inverse Function Theorem:

**(B3')** *There are neighborhoods $u \subseteq [0,1]\times[0,1]$ of $p$ and $U \subseteq \pi([0,1]\times[0,1])$ of $P$ such that $\pi(x) \in U$ if and only if $x \in u$, and $\pi|_u$ is a diffeomorphism.*

This statement implies the obvious analog of (2.1) for $p_i \in I\!R^2$ and $P_i \in I\!R^5$.

We begin this case, as before, by examining the Taylor expansions of $\pi$ and $\pi \circ f$ about a point $P = \pi(p)$ in $I\!R^5$ for which (B3) and (B4) hold. With $h = (h_1, h_2) \in I\!R^2$, any point $P + \Delta P$ in $\pi(u)$ near $P$ can be written in the form

$$\begin{aligned} P + \Delta P &= \pi(p+h) \\ &= P + h_1\pi_x(p) + h_2\pi_y(p) + \frac{1}{2}h_1^2\pi_{xx}(p) \\ &\quad + h_1h_2\pi_{xy}(p) + \frac{1}{2}h_2^2\pi_{yy}(p) + Rem(h,p) \end{aligned}$$

where $Rem(h,p)$ is the Taylor remainder term, now a vector in $I\!R^5$. Again, there is a constant $C_{Taylor} > 0$ such that $\|Rem(h,p)\| \le C_{Taylor}\|h\|^3$ for sufficiently small $h$. The **canonical embedding basis** (for $I\!R^5$) at $P$ will consist of the five derivative vectors from (B4): $\pi_x(p)$, $\pi_y(p)$, $\pi_{xx}(p)$, $\pi_{xy}(p)$, and $\pi_{yy}(p)$. With respect to this basis, we can write

$$\Delta P = \left(h_1, h_2, \frac{1}{2}h_1^2, h_1h_2, \frac{1}{2}h_2^2\right)_P + Rem(h,p). \tag{2.5}$$

Next, look at the image $F(P) = \pi(f(p))$. The Taylor expansion of $F(P+\Delta P) = F(\pi(p+h)) = \pi(f(p+h))$ is given by

$$\begin{aligned} F(P+\Delta P) &= F(P) + h_1(\pi\circ f)_x(p) + h_2(\pi\circ f)_y(p) + \frac{1}{2}h_1^2(\pi\circ f)_{xx}(p) \\ &\quad + h_1h_2(\pi\circ f)_{xy}(p) + \frac{1}{2}h_2^2(\pi\circ f)_{yy}(p) + Rem_f(h,p) \end{aligned}$$

Here, $Rem_f$ is the Taylor remainder vector associated with $\pi \circ f$, and without loss of generality, it too satisfies $\|Rem_f(h,p)\| \le C_{Taylor}\|h\|^3$ for small $h$.

**Definition 2.3.** *Assuming (B1) – (B4), we define the* **Eckmann-Ruelle linearization** *$M = M(P)$ to be the unique linear map on $I\!R^5$ satisfying*

$$\begin{aligned} M\pi_x(p) &= (\pi\circ f)_x(p) & \qquad M\pi_{xx}(p) &= (\pi\circ f)_{xx}(p) \\ M\pi_y(p) &= (\pi\circ f)_y(p) & M\pi_{xy}(p) &= (\pi\circ f)_{xy}(p) \\ & & M\pi_{yy}(p) &= (\pi\circ f)_{yy}(p) \end{aligned} \tag{2.6}$$

We will prove in Theorem 2.6 that this linear map $M(P)$ provides the best local linearization of $F$ in $I\!R^5$ about the base point $P$. The proof of Theorem 2.6 will be virtually identical to that of Theorem 2.2 whose main idea was the systematic cancellation of low order terms between the Taylor expansions about

a point and its image. Up to second order, there are five distinct terms to be eliminated, each requiring a basis element. Thus, we embed into $\mathbb{R}^5$. Likewise, to kill off all third-order terms in a longer expansion, we would need a total of nine terms, and so we would embed in $\mathbb{R}^9$. This is more restrictive than the one-dimensional case above since we cannot embed into an arbitrary $\mathbb{R}^m$.

Theorem 2.6 requires a slightly stronger notion of limit point. Since the underlying phase space has more than one dimension, we should be able to approach any underlying base point from more than two directions. This leads to the definition of an "approach direction." The idea is that there is a line extending out from the base point and some subsequence of points from the trajectory approach the base point along this line.

**Definition 2.4.** *Let $l$ be a unit vector in $\mathbb{R}^m$. A subset $S$ of $\mathbb{R}^m$ has the* **approach direction** *$l$ at the base point $P \in \mathbb{R}^m$ provided there is a sequence $\{Q_k\}_{k=1}^{\infty}$ from $S$ such that:*

1. *$Q_k \to P$ as $k \to \infty$, and*
2. *$\frac{\Delta Q_k}{\|\Delta Q_k\|} \to l$ as $k \to \infty$, where $\Delta Q_k := Q_k - P$.*

*We call a collection of approach directions at $P$* **distinct** *provided that no two are the same and no two are reflections through the origin. Multiple approach directions are not required to be linearly independent.*

Figure 2.2 shows three distinct approach directions at a point $P$ on a fractal attractor in $\mathbb{R}^2$. One can find trajectory points converging to $P$ which are on or near the intersections of the lines $l_1$, $l_2$, and $l_3$ with the attractor. This behavior was automatic in the one-dimensional case, and points in hyperbolic systems will have multiple approach directions. It may even be the case that typical points in arbitrary (multi-dimensional) chaotic systems have infinitely-many distinct approach directions.

It is important to be able to relate the (often fractal) geometry of the reconstructed attractor back to the geometry of the underlying attractor. One reason for this, of course, is that we can only observe the geometry of the reconstruction. Another reason is that we need access to facts about the underlying system in order to extract information from the observations. These reasons provide motivation for the next lemma whose proof is straightforward when $\pi$ is a diffeomorphism of a neighborhood of $P$.

**Lemma 2.5.** *Assume (B1) and (B2). Let $P = \pi(p)$ be a point of $\pi([0,1] \times [0,1])$ for which (B3) and (B4) hold. If $P$ is in the closure of a trajectory of $F$, $P_0, P_1, \cdots \in \mathbb{R}^5$ that has $d$ distinct approach directions at $P$, then the underlying trajectory $p_0, p_1, \cdots \in \mathbb{R}^2$, where $P_i = \pi(p_i)$, has $d$ distinct approach directions at the point $p \in [0,1] \times [0,1]$.*

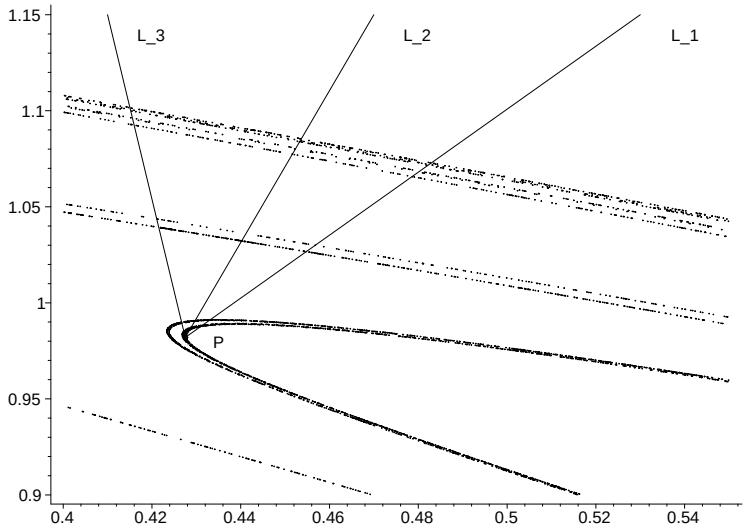



Figure 2.2: The trajectory has three approach directions $L_1$, $L_2$, and $L_3$ at the base point $p$. The trajectory is dense in the fractal attractor shown, so data points are available arbitrarily close to the intersections of the three lines with the attractor.

We prove the Local Linearization Theorem for the case where we reconstruct a two-dimensional underlying system in $\mathbb{R}^5$. In the statement below, we require the reconstructed trajectory in $\mathbb{R}^5$ to have three distinct approach directions at the point $P$. We need three distinct approach directions in order to determine the Eckmann-Ruelle linearization. Suppose, for instance, that the available data points approach the base point $P \in \mathbb{R}^5$ only along two directions. By Lemma 2.5, the underlying attractor in $\mathbb{R}^2$ will also have two distinct approach directions. Suppose that the data lies exactly on the $x$- and $y$-axes (in local coordinates in $\mathbb{R}^2$). We can compute $\pi_x(p)$, $\pi_{xx}(p)$, ... from points of the form $(h_1, 0)$. Likewise, we can compute $\pi_y(p)$, $\pi_{yy}(p)$, ... from points of the form $(0, h_2)$. However, at each of these points, all terms involving mixed partial derivatives vanish from the Taylor expansions, making it impossible to compute $\pi_{xy}(p)$. Of course, without knowledge of $\pi_{xy}(p)$ (and where it maps), we cannot uniquely determine a best local linearization at $P$. This is why we require three distinct approach directions.

**Theorem 2.6 (Local Linearization, $\mathbb{R}^2 \to \mathbb{R}^5$).** *Assume (B1) and (B2). Let $P = \pi(p)$ be a point of $\pi([0,1] \times [0,1])$ for which (B3) and (B4) hold. If $P$ is in the closure of a trajectory of $F$, $P_0, P_1, \cdots \in \mathbb{R}^5$ that has three distinct approach directions at $P$, then the Eckmann-Ruelle linearization $M = M(P)$ defined by (2.6) is the unique linear map such that*

$$F(P + \Delta P) - F(P) - M\Delta P = O\Big(\|\Delta P\|^3\Big) \qquad (2.7)$$
$$\text{as } \|\Delta P\| \to 0, \qquad \text{where } P + \Delta P \in \pi([0,1] \times [0,1])$$

**Proof.** We use the same notation as in the proof of Theorem 2.2, except that we write $h = (h_1, h_2) \in I\!R^2$. The first thing to do is show that the Eckmann-Ruelle linearization $M = M(P)$ satisfies (2.7). As in the proof of Theorem 2.2, we use the definition of $M$ in equation (2.6) to cancel terms from the Taylor expansions. Thus,

$$\begin{aligned} \|F(P+\Delta P) - F(P) - M\Delta P\| &= \|Rem_f(h,p) - MRem(h,p)\| \\ &\le C_{Taylor}(1 + \|M\|)\,\|h\|^3 \end{aligned}$$

for small $h$. Therefore, by (2.1), for small $\Delta P$ with $P + \Delta P \in \pi([0,1]\times[0,1])$,

$$\|F(P+\Delta P) - F(P) - M\Delta P\| \le C_{Taylor}(1 + \|M\|)C_\pi^3\,\|\Delta P\|^3\,,$$

establishing (2.7).

It remains to show that $M$ is the only matrix that satisfies (2.7). Suppose to the contrary there is another matrix $\widetilde{M}$ such that

$$\left\|F(P+\Delta P) - F(P) - \widetilde{M}\Delta P\right\| \le C_1\|\Delta P\|^3 \qquad \text{as } \Delta P \to 0,\ P+\Delta P \in \pi(u)$$

for some constant $C_1$. Set $A := \widetilde{M} - M$. For small $\Delta P$ with $P + \Delta P \in \pi(u)$,

$$\begin{aligned} \|A\Delta P\| &\le \|F(P+\Delta P) - F(P) - M\Delta P\| + \left\|F(P+\Delta P) - F(P) - \widetilde{M}\Delta P\right\| \\ &\le C_2\|\Delta P\|^3. \end{aligned}$$

With respect to the canonical embedding basis at $P$, the small vectors $\Delta P$ have the form shown in (2.5), and so

$$\begin{aligned} \left\|A\left(h_1, h_2, \frac{1}{2}h_1^2, h_1h_2, \frac{1}{2}h_2^2\right)_P\right\| &\le \|A\Delta P\| + \|ARem(h,p)\| \\ &\le C_2\|\Delta P\|^3 + \|A\|\,\|Rem(h,p)\| \\ &\le C_2C_\pi^3\,\|h\|^3 + \|A\|C_{Taylor}\,\|h\|^3 \\ &:= C_3\,\|h\|^3 \end{aligned}$$

Thus, for all tangent vectors $h \in u$ sufficiently small

$$\left\|A\left(h_1, h_2, \frac{1}{2}h_1^2, h_1h_2, \frac{1}{2}h_2^2\right)_P\right\|^2 \le C_3^2\,\|h\|^6\,.$$

If we represent the matrix $A$ with respect to the canonical embedding basis at $P$, then the left-hand side is a polynomial $p(h_1, h_2)$ with degree at most 4. At last, we use the assumption that we have three distinct approach directions. It follows from Lemma 2.5, that the underlying trajectory $p_0, p_1, \cdots \in I\!R^2$ also has three distinct approach directions. This fact, together with the inequality above, gives us precisely the hypotheses needed to apply Proposition A.3 of Appendix A.

Thus, $p(h_1, h_2)$ must be identically zero, and it follows that all of the coefficients of the matrix $A$ must also be zero. Therefore, $\widetilde{M} = M(P)$ proving that $M(P)$ is indeed the unique matrix satisfying (2.7). □

This proof extends further to the case when we embed a two-dimensional system into higher dimensions. To do this, we first need to embed into an appropriate dimension. In the proof, we cancelled all of the first-order and second-order terms in a Taylor expansion. In the general setting, we want to cancel all terms of order up to and including order $D$. There are $\sum_{d=1}^{D} \binom{d+1}{1} = \frac{1}{2}D(D+3)$ such terms. For each term to be cancelled, we need a basis vector in the canonical embedding basis. Therefore, we can embed our two-dimensional system into any $m$-dimensional Euclidean space where $m = \frac{1}{2}D(D+3)$ for some $D > 0$. For $D = 2$ we have $m = 5$, and for $D = 3$ we have $m = 9$.

We now assume that we are given some $D > 1$ with $m = \frac{1}{2}D(D+3)$ the corresponding dimension. We make the following specific assumptions about $f$, $\pi$, and a point $P = \pi(p) \in I\!R^m$.

**(B5)** *$f$ maps the unit square $[0,1] \times [0,1]$ into itself, and $f$ is $C^{D+1}$.*

**(B6)** *$\pi$ maps $[0,1] \times [0,1]$ into $I\!R^m$, and $\pi$ is $C^{D+1}$.*

**(B7)** *For this $p \in [0,1] \times [0,1]$, $P = \pi(p) \in I\!R^m$, and the set $\pi^{-1}(P)$ contains only one point, namely $p$.*

**(B8)** *The various partial derivative vectors for $\pi$ at $p$ of order at most $D$, namely*

$$\begin{array}{lllllll} \pi_x(p) & := & \frac{\partial \pi}{\partial x}(p) & \pi_{xx}(p) & := & \frac{\partial^2 \pi}{\partial x^2}(p) & \\ \pi_y(p) & := & \frac{\partial \pi}{\partial y}(p) & \pi_{xy}(p) & := & \frac{\partial^2 \pi}{\partial x \partial y}(p) & \cdots \\ & & & \pi_{yy}(p) & := & \frac{\partial^2 \pi}{\partial^2 y}(p) & \end{array} \qquad \begin{array}{lll} \pi_{xx\ldots xx}(p) & := & \frac{\partial^D \pi}{\partial x^D}(p) \\ \pi_{xx\ldots xy}(p) & := & \frac{\partial^D \pi}{\partial x^{D-1} \partial y}(p) \\ & \vdots & \\ \pi_{yy\ldots yy}(p) & := & \frac{\partial^D \pi}{\partial y^D}(p) \end{array}$$

*are linearly independent in $I\!R^m$.*

Properties (B7) and (B8) are generic properties in the space of $C^{D+1}$ functions from $[0,1] \times [0,1]$ into $I\!R^m$. The $m$ vectors in (B8) form the canonical embedding basis at $P$. As before, we define the **Eckmann-Ruelle linearization** at $P$ to be the unique linear map $M = M(P)$ defined by the relations :

$$\begin{array}{lll} M\pi_x(p) & = & (\pi \circ f)_x(p) \\ M\pi_y(p) & = & (\pi \circ f)_y(p) \end{array} \qquad \begin{array}{lll} M\pi_{xx}(p) & = & (\pi \circ f)_{xx}(p) \\ M\pi_{xy}(p) & = & (\pi \circ f)_{xy}(p) \\ M\pi_{yy}(p) & = & (\pi \circ f)_{yy}(p) \end{array}$$

$$\begin{array}{rrcl} & M\pi_{xx\ldots xx}(p) & = & (\pi\circ f)_{xx\ldots xx}(p) \\ & M\pi_{xx\ldots xy}(p) & = & (\pi\circ f)_{xx\ldots xy}(p) \\ \cdots & & \vdots & \\ & M\pi_{yy\ldots yy}(p) & = & (\pi\circ f)_{yy\ldots yy}(p) \end{array} \tag{2.8}$$

Here is the general theorem for two-dimensional underlying dynamics.

**Theorem 2.7 (Local Linearization, $\mathbb{R}^2 \to \mathbb{R}^m$).** *Let $m = \frac{1}{2}D(D+3)$ where $D > 1$. Assume (B5) and (B6). Let $P = \pi(p)$ be a point of $\pi([0,1]\times[0,1])$ for which (B7) and (B8) hold. If $P$ is in the closure of a trajectory of $F$, $P_0, P_1, \cdots \in \mathbb{R}^m$, and if the trajectory has $D+1$ distinct approach directions at $P$, then the Eckmann-Ruelle linearization $M = M(P)$ defined by (2.8) is the unique linear map such that*

$$\begin{array}{rcl} F(P+\Delta P) - F(P) - M\Delta P & = & O\Big(\|\Delta P\|^{D+1}\Big) \\ \text{as } \|\Delta P\| \to 0, & & \text{where } P+\Delta P \in \pi([0,1]\times[0,1]) \end{array} \tag{2.9}$$

The proof of this is virtually identical to the proof of Theorem 2.6 and is omitted. Note that the polynomial proposition A.3 is already stated generally enough to be used for this proof.

The Eckmann-Ruelle linearization will also exist in certain other cases, such as when reconstructing a three-dimensional underlying system into $\mathbb{R}^9$. The arguments above extend in a natural way to certain cases where we have a dynamical system $f$ on $\mathbb{R}^n$ being reconstructed in a larger dimensional space $\mathbb{R}^m$. Assume that $D > 1$ and $f$ and $\pi$ each have at least $D+1$ continuous derivatives in each coordinate. Note that there are $\sum_{d=1}^{D}\binom{d+n-1}{n-1} = \binom{D+n}{n} - 1$ terms of order at most $D$ in the Taylor series of $f$ and $\pi$. Then, when $m = \binom{D+n}{n} - 1$, we can construct the canonical embedding basis as in assumption (B8) and define the Eckmann-Ruelle linearization $M(P)$ as in (2.8). Because (2.8) guarantees that the Taylor series will collapse nicely, it is easy to see that $M(P)$ satisfies (2.9). We will also need some mild geometric condition similar to those given in the previous theorems to guarantee that $M(P)$ is the only matrix satisfying (2.9). We will not pursue these ideas further at this time.

# Chapter 3

# Lyapunov Exponent Formulas

Once the local linearization matrices have been computed at each point of the reconstructed trajectory, we must extract Lyapunov exponents from them. Mimicking the standard definition of Lyapunov exponents (see for example, [12, pp. 31]), we define the **Eckmann-Ruelle-Lyapunov (ERL) exponents** of the reconstructed trajectory $P_0, P_1, \ldots$ in $\mathbb{R}^m$ to be the values obtained by the limit

$$h_{ER}(P_0, \nu) := \lim_{n\to\infty} \frac{1}{n} \ln \|M_{n-1} M_{n-2} \cdots M_1 M_0 \nu\| \tag{3.1}$$

for unit vectors $\nu \in \mathbb{R}^m$, where $M_i = M(P_i)$ is the best local linearization (i.e., Eckmann-Ruelle linearization) at the point $P_i = \pi(p_i)$ in the trajectory. We need a separate definition here because the standard definition of Lyapunov exponent uses the Jacobian derivative, which need not exist along the trajectory. In this chapter, we show that the matrix product in (3.1) can be written as an upper triangular matrix. A straightforward calculation will then produce a formula for the limiting values of (3.1).

In practice, the limit in (3.1) can be computed using the treppen-iteration algorithm described in [1, 2] and elsewhere. This method uses QR matrix decomposition to convert the matrix product $M_{n-1} \cdots M_0$ into a product of upper triangular matrices $R_{n-1} \cdots R_0$. The diagonal elements of the latter upper triangular matrix are the products of the corresponding diagonal elements of the $R_i$. Then, we can read the Lyapunov exponents right from the diagonals of the intermediate matrices $R_i$:

$$\lambda_j = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln \left| (R_i)_{jj} \right| .$$

Theorem 3.2 justifies the ability to read the exponents directly from the diagonal in this way. A proof of the theorem will be given in Appendix B.

**Definition 3.1.** *A sequence of real numbers* $\{r_n\}$ *has* **(geometric) growth rate** $\gamma$ *provided*

$$\lim_{n\to\infty} \frac{\ln |r_n|}{n} = \gamma.$$

**Theorem 3.2.** *For $k = 1, 2 \ldots$, let $A_k$ be $m \times m$ upper triangular matrices, and define $S_n = A_n \cdots A_1$. Assume the magnitudes of the entries of $A_k$ are bounded independent of $k$, and that the diagonal entries of $S_n$ have growth rates $\gamma_1, \ldots, \gamma_m$ as $n \to \infty$. Then there exist vectors $v_1, \ldots, v_m \in \mathbb{R}^m$ such that for each $i = 1, \ldots, m$, $\|S_n v_i\|$ has growth rate $\gamma_i$.*

The next theorem and its two-dimensional analog, Theorem 3.6, are the main results of this paper. In both theorems, most of the hypotheses are used to guarantee that the Eckmann-Ruelle linearization exists at each point of the trajectory.

**Theorem 3.3 (Lyapunov Exponent Formula, $\mathbb{R}^1 \to \mathbb{R}^m$).** *Assume (A1), (A2) and that the trajectory of $f$ in $[0, 1]$, $p_0, p_1, \ldots$, has Lyapunov exponent $\lambda$. If each point of the reconstructed trajectory, $P_i = \pi(p_i) \in \mathbb{R}^m$, satisfies (A3) and (A4) and is a limit point of the trajectory, then the reconstructed trajectory has Eckmann-Ruelle-Lyapunov exponents $\lambda, 2\lambda, 3\lambda, \ldots, m\lambda$.*

In order to prove this, a few lemmas will be useful. Lemma 3.4 is a technical lemma leading to the matrix representation for $M(P)$ given in Lemma 3.5. With this matrix representation, we will prove Theorem 3.3.

**Lemma 3.4.** *Let $J$ be an interval, and let $k, d$ be positive integers. If $f : J \to J$ is a $C^{k+1}$ function, then for $1 \le j \le i \le k$ there are differentiable scalar functions $a_{ij}(t)$, such that for any $C^{k+1}$ function $\phi : J \to \mathbb{R}^d$, $\phi(t) = (\phi_1(t), \ldots, \phi_d(t))$, we have*

$$(\phi \circ f)^{(i)}(t) = \sum_{j=1}^{i} a_{ij}(t)\phi^{(j)}(f(t)), \qquad \text{for } i = 1, \ldots, k \tag{3.2}$$

*where $\phi^{(j)}(t) = \left(\frac{d^j \phi_1}{dt^j}(t), \ldots, \frac{d^j \phi_d}{dt^j}(t)\right)$ represents the $j$-th derivative vector of $\phi(t)$.*

Note that the scalar functions $a_{ij}(t)$ depend only on $f$ and not on the other function $\phi$. As will be seen in the proof, these functions $a_{ij}(t)$ are the results of collecting terms involving the derivatives of $\phi$.

**Proof of Lemma 3.4.** Since vector-valued functions of a real-variable are naturally differentiated componentwise, it is enough to prove the lemma for any one component. Thus, we may assume $d = 1$ and $\phi : J \to \mathbb{R}$.

The proof is by induction. For $i = 1$, we apply the Chain Rule:

$$(\phi \circ f)^{(1)}(t) = f'(t)\phi^{(1)}(f(t))$$

and note that $a_{11}(t) := f'(t)$ is $C^k$. For $i = 2$, we apply the chain rule again:

$$(\phi \circ f)^{(2)}(t) = f'(t)^2 \phi^{(2)}(f(t)) + f''(t)\phi^{(1)}(f(t)).$$

Note that $a_{21}(t) := f''(t)$ and $a_{22}(t) := (f'(t))^2$ are both $C^{k-1}$.

For induction, suppose that (3.2) holds for $i$ where the $a_{ij}(t)$ are $C^{k+1-i}$, depending only on $f$, not on $\phi$. We differentiate (3.2) and collect together terms involving the derivatives of $\phi$. Specifically, if we define the functions $a_{(i+1)j}(t)$ using

$$\begin{aligned} a_{(i+1)1}(t) &:= a'_{i1}(t) \\ a_{(i+1)j}(t) &:= a'_{ij}(t) + a_{i(j-1)}(t)f'(t), \qquad j = 2, \ldots, i-1 \\ a_{(i+1)i}(t) &:= if'(t)^{i-1}f''(t) + a_{i(i-1)}(t)f'(t) \\ a_{(i+1)(i+1)}(t) &:= f'(t)^{i+1} \end{aligned} \tag{3.3}$$

then we obtain from differentiating (3.2):

$$\begin{aligned} (\phi \circ f)^{(i+1)}(t) &= \frac{d}{dt}\left( f'(t)^i \phi^{(i)}(f(t)) + \sum_{j=1}^{i-1} a_{ij}(t)\phi^{(j)}(f(t)) \right) \\ &= if'(t)^{i-1}f''(t)\phi^{(i)}(f(t)) + f'(t)^{i+1}\phi^{(i+1)}(f(t)) \\ &\quad + \sum_{j=1}^{i-1} \left( a'_{ij}(t)\phi^{(j)}(f(t)) + a_{ij}(t)f'(t)\phi^{(j+1)}(f(t)) \right) \\ &= \sum_{j=1}^{i+1} a_{(i+1)j}(t)\phi^{(j)}(f(t)). \end{aligned}$$

Note that the functions $a_{(i+1)j}(t)$ are $C^{k+1-(i+1)}$ and that they depend only on the function $f$, not on the other function $\phi$. This completes the proof. □

**Lemma 3.5.** *Assume (A1) and (A2). Let $P = \pi(p) \in \mathbb{R}^m$ and let $Q = F(P)$ be points satisfying (A3) and (A4) in the closure of a trajectory of $F$. Let $\alpha$ and $\beta$ represent the canonical embedding bases at $P$ and $Q$, respectively. With respect to these bases, the Eckmann-Ruelle linearization $[M(P)]_\alpha^\beta$ is upper triangular with diagonal elements $M(P)_{jj} = f'(p)^j$.*

**Proof.** Recall that the canonical embedding basis at a point $P = \pi(p)$ in $\mathbb{R}^m$ consists of the vectors $\pi^{(1)}(p), \ldots, \pi^{(m)}(p)$ in $\mathbb{R}^m$. Set $M = M(P)$ and $q = f(p)$. By the definition of the Eckmann-Ruelle linearization (2.3), we have $M\pi^{(i)}(p) = (\pi \circ f)^{(i)}(p)$. We must write the vectors $(\pi \circ f)^{(i)}(p)$ in terms of the canonical embedding basis at $Q = \pi(q)$. Applying Lemma 3.4 with $k = d = m$ and $\phi = \pi$, we have

$$M\pi^{(i)}(p) = (\pi \circ f)^{(i)}(p) = f'(p)^i \pi^{(i)}(q) + \sum_{j=1}^{i-1} a_{ij}(p)\pi^{(j)}(q) \qquad \text{for } i = 1, \ldots, m$$

because $f(p) = q$. This provides a representation for $M$ with respect to the canonical embedding bases $\alpha$ and $\beta$:

$$[M]_\alpha^\beta = \begin{pmatrix} f'(p) & a_{21}(p) & a_{31}(p) & \cdots & a_{k1}(p) \\ 0 & f'(p)^2 & a_{32}(p) & \cdots & a_{k2}(p) \\ 0 & 0 & f'(p)^3 & \cdots & a_{k3}(p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f'(p)^m \end{pmatrix}.$$

□

It is important to note that this matrix form for the Eckmann-Ruelle linearization is completely independent of the embedding $\pi$. Information about $\pi$ is used to form the coordinate system with respect to which we view the dynamics; however, once inside that coordinate system, we only see the action of $f$.

**Proof of Theorem 3.3.** For each point $P_i \in I\!R^m$, the Eckmann-Ruelle linearization $M_i = M(P_i)$ exists. To compute the ERL exponents, we must evaluate the limit (3.1). Let $\beta_i$ denote the canonical embedding basis at $P_i$. Notice that the canonical embedding bases for the upper triangular matrix representation fit together perfectly:

$$[M_{i+1}]_{\beta_{i+1}}^{\beta_{i+2}} [M_i]_{\beta_i}^{\beta_{i+1}} = [M_{i+1} M_i]_{\beta_i}^{\beta_{i+2}}.$$

It follows that the matrix product $M_{n-1} M_{n-2} \cdots M_1 M_0$ is upper triangular when expressed with respect to the canonical embedding bases $\beta_0$ and $\beta_n$. That is, the matrix product $[M_{n-1} M_{n-2} \cdots M_1 M_0]_{\beta_0}^{\beta_n}$ can be written

$$\begin{pmatrix} \prod_{j=0}^{n-1} f'(p_j) & b_{12} & b_{13} & \cdots & b_{1k} \\ 0 & \left(\prod_{j=0}^{n-1} f'(p_j)\right)^2 & b_{23} & \cdots & b_{2k} \\ 0 & 0 & \left(\prod_{j=0}^{n-1} f'(p_j)\right)^3 & \cdots & b_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \left(\prod_{j=0}^{n-1} f'(p_j)\right)^m \end{pmatrix}$$

where the $b_{ij}$ are numbers which depend solely on the underlying dynamical system $f$ and the first $n$ points of the trajectory of $p_0$ in $[0, 1]$.

We apply Theorem 3.2 to complete the proof. Note that the elements of each $M_i$ (with respect to the appropriate coordinate systems) are combinations of the

first $m$ derivatives of $f$, each of which is continuous. These elements will be bounded independent of $i$ on any compact set containing the entire trajectory. By hypothesis, the Lyapunov exponent $\lambda$ of $f$ is given by

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \ln \left| \prod_{j=0}^{n-1} f'(p_j) \right|,$$

and it is clear that the diagonal entries of the matrix product $M_{n-1} \cdots M_0$ have growth rates $\lambda, 2\lambda, \ldots, m\lambda$. Theorem 3.2 now provides the vectors which grow at these characteristic rates, and we conclude that the ERL exponents are $\lambda, 2\lambda, \ldots, m\lambda$. □

With this theorem, we expect that whenever the underlying system is one-dimensional with Lyapunov exponent $\lambda$, we will compute exponents $\lambda, 2\lambda, \ldots, m\lambda$ in the absence of noise. Note that if the system has a positive Lyapunov exponent (and hence is chaotic), then the true exponent will be the *smallest* of the computed numbers!

Next, we consider the case where the underlying system is two-dimensional and reconstructed in $\mathbb{R}^5$. The basic arguments are similar though the situation is more involved, as we shall see. The formula for the computed Lyapunov exponents in this case is given in the next theorem.

**Theorem 3.6 (Lyapunov Exponent Formula, $\mathbb{R}^2 \to \mathbb{R}^5$).** *Assume (B1), (B2), and that the trajectory of $f$ in $[0,1] \times [0,1]$, $p_0, p_1, \ldots,$ has Lyapunov exponents $\lambda$ and $\mu$. Assume each point of the reconstructed trajectory, $P_i = \pi(p_i)$ in $\mathbb{R}^5$, satisfies (B3) and (B4), and the set $\{P_0, P_1, P_2, \ldots\}$ in $\mathbb{R}^5$ has at least three distinct approach directions at each $P_i$. Then, the reconstructed trajectory has Eckmann-Ruelle-Lyapunov exponents $\lambda$, $\mu$, $2\lambda$, $\lambda + \mu$, and $2\mu$.*

**Proof.** Recall that the canonical embedding basis given in assumption (B4) from Chapter 2 consists of the first and second order partial derivative vectors of $\pi$, namely $\pi_x(p)$, $\pi_y(p)$, $\pi_{xx}(p)$, $\pi_{xy}(p)$, and $\pi_{yy}(p)$. In fact, the construction of the Eckmann-Ruelle linearization from the Taylor series of $f$ and $\pi$ can be carried out with respect to any orthonormal set of coordinates. The uniqueness part of the Local Linearization Theorem 2.6 ensures that the resulting linear map will be the same. Thus, we may introduce convenient local coordinate systems at each point $p_i \in \mathbb{R}^2$ of the trajectory of $p_0$.

Without loss of generality, we assume $\lambda \geq \mu$. We begin by choosing a unit Lyapunov vector $v_0 \in \mathbb{R}^2$ corresponding to the exponent $\mu$ at the initial point $p_0 \in \mathbb{R}^2$ of the underlying trajectory. That is, we choose a unit vector $v_0$ such that

$$\lim_{n\to\infty} \frac{1}{n} \ln \left\| Df_{p_{n-1}} Df_{p_{n-2}} \cdots Df_{p_1} Df_{p_0} v_0 \right\| = \mu.$$

By Oseledec's Multiplicative Ergodic Theorem [1], almost every other vector in $\mathbb{R}^2$ has growth rate $\lambda$. Choose a unit vector $w_0$ perpendicular to $v_0$. This gives us an orthonormal basis $\{v_0, w_0\}$ for $\mathbb{R}^2$ based at $p_0$. In general, given the basis $\{v_n, w_n\}$ at $p_n \in \mathbb{R}^2$, we construct the basis at $p_{n+1}$ by setting $v_{n+1} := Df_{p_n} v_n / \|Df_{p_n} v_n\|$ and choosing $w_{n+1}$ to be the unit vector perpendicular to $v_{n+1}$ that satisfies $\langle w_{n+1}, Df_{p_n} w_n \rangle > 0$. Thus, at each point $p_n$, we have an orthonormal basis for $\mathbb{R}^2$. For each $n$, we write points $p' \in \mathbb{R}^2$ near $p_n$ as $p' = (x, y)$ provided $p' = p_n + xv_n + yw_n$. Now, we can describe the underlying dynamics $f$ near $p_n$ in terms of these bases at $p_n$ and $p_{n+1}$ by $f(x, y) = (g(x, y), h(x, y))$. Though we will not explicitly show it, this representation for $f$ depends on the base point $p_n$, and the reader should keep in mind that the component functions $g$ and $h$ may look very different as we vary $p_n$. Note that $h_x(p_n) = 0$ because $Df_{p_n} v_n = \|Df_{p_n} v_n\| v_{n+1}$ has no $y$-component at $p_{n+1}$. Thus, we can write the Jacobian derivative of $f$ at $p_n$ as:

$$Df_{p_n} = \begin{pmatrix} g_x(p_n) & g_y(p_n) \\ 0 & h_y(p_n) \end{pmatrix}.$$

Following the ideas in the previous case, a calculation using the chain rule produces the next set of equations as the analog of (3.2), where we write $q_n = f(p_n)$ and the partial derivatives of $g$ and $h$ are evaluated at $p_n$:

$$\begin{aligned}
(\pi \circ f)_x(p_n) &= g_x \pi_x(q_n) \\
(\pi \circ f)_y(p_n) &= g_y \pi_x(q_n) + h_y \pi_y(q_n) \\
(\pi \circ f)_{xx}(p_n) &= g_{xx} \pi_x(q_n) + h_{xx} \pi_y(q_n) + g_x^2 \pi_{xx}(q_n) \\
(\pi \circ f)_{xy}(p_n) &= g_{xy} \pi_x(q_n) + h_{xy} \pi_y(q_n) + g_x g_y \pi_{xx}(q_n) + g_x h_y \pi_{xy}(q_n) \\
(\pi \circ f)_{yy}(p_n) &= g_{yy} \pi_x(q_n) + h_{yy} \pi_y(q_n) + g_y^2 \pi_{xx}(q_n) + 2 g_y h_y \pi_{xy}(q_n) + h_y^2 \pi_{yy}(q_n)
\end{aligned}$$

For each $n$, let $\beta_n$ denote the canonical embedding basis for $\mathbb{R}^5$ at $P_n = \pi(p_n)$. Representing the Eckmann-Ruelle linearization with respect to $\beta_n$ and $\beta_{n+1}$ as we did in Lemma 3.5, we obtain this matrix representation for $M_n = M(P_n)$:

$$[M_n]_{\beta_n}^{\beta_{n+1}} = \begin{pmatrix} g_x & g_y & g_{xx} & g_{xy} & g_{yy} \\ 0 & h_y & h_{xx} & h_{xy} & h_{yy} \\ 0 & 0 & g_x^2 & g_x g_y & g_y^2 \\ 0 & 0 & 0 & g_x h_y & 2 g_y h_y \\ 0 & 0 & 0 & 0 & h_y^2 \end{pmatrix}$$

All of the terms above the diagonal come from combinations of first and second derivatives of $f$, and, because $f$ is $C^3$ by hypothesis, they will be bounded independent of $p_n$. Note that the upper left $2 \times 2$ block is the Jacobian $Df_{p_n}$ of the underlying dynamics and that the lower $3 \times 3$ block contains only combinations of terms from $Df_{p_n}$.

As in the proof of Theorem 3.3, the product $[M_{n-1}\cdots M_0]_{\beta_0}^{\beta_n}$ is upper triangular when written with respect to the canonical embedding bases:

$$\begin{pmatrix} \prod_{j=0}^{n-1} g_x(p_j) & b_{12} & b_{13} & b_{14} & b_{15} \\ 0 & \prod_{j=0}^{n-1} h_y(p_j) & b_{23} & b_{24} & b_{25} \\ 0 & 0 & \left(\prod_{j=0}^{n-1} g_x(p_j)\right)^2 & b_{34} & b_{35} \\ 0 & 0 & 0 & \prod_{j=0}^{n-1} g_x(p_j)h_y(p_j) & b_{45} \\ 0 & 0 & 0 & 0 & \left(\prod_{j=0}^{n-1} h_y(p_j)\right)^2 \end{pmatrix}$$

where the $b_{ij}$ are numbers which depend solely on the underlying dynamical system $f$ and the first $n$ points of the trajectory of $p_0$ in $[0,1]\times[0,1]$. Now, we need to compute the growth rates of the diagonal terms. To do this, we compute the Lyapunov exponents $\lambda$ and $\mu$ of the underlying dynamical system in terms of the components of the Jacobian matrices. Since $Df_{p_n}v_n = \|Df_{p_n}v_n\|v_{n+1}$, we have $g_x(p_n) = \|Df_{p_n}v_n\|$. It follows that

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\ln\left(\prod_{i=0}^{n-1} g_x(p_n)\right) &= \lim_{n\to\infty}\frac{1}{n}\ln\left(\prod_{i=0}^{n-1}\|Df_{p_i}v_i\|\right) \\ &= \lim_{n\to\infty}\frac{1}{n}\ln\|Df_{p_{n1}}\cdots Df_{p_0}v_0\| \\ &= \mu \end{aligned} \tag{3.4}$$

because $v_0$ was chosen to be a Lyapunov vector for $\mu$. Recall from [1, pp. 632] that the growth rate of areas is given by the sum of the Lyapunov exponents. Thus,

$$\begin{aligned} \mu+\lambda &= \lim_{n\to\infty}\frac{1}{n}\ln\left|\det\left(Df^n{}_{p_0}\right)\right| \\ &= \lim_{n\to\infty}\frac{1}{n}\ln\left(\prod_{i=0}^{n-1}\left|\det\left(Df_{p_i}\right)\right|\right) \\ &= \lim_{n\to\infty}\frac{1}{n}\ln\left(\prod_{i=0}^{n-1}\left|g_x(p_n)h_y(p_n)\right|\right) \\ &= \mu + \lim_{n\to\infty}\frac{1}{n}\ln\left(\prod_{i=0}^{n-1}\left|h_y(p_n)\right|\right) \end{aligned}$$

and it follows that

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \ln \left( \prod_{i=0}^{n-1} |h_y(p_n)| \right). \tag{3.5}$$

Finally, we see that the diagonal terms have growth rates of $\lambda$, $\mu$, $2\lambda$, $\lambda + \mu$, and $2\mu$. It follows from Theorem 3.2 that the reconstructed trajectory, $P_0, P_1, \ldots$ in $I\!R^5$ has these ERL exponents, completing the proof of Theorem 3.6. □

# Chapter 4

# Convergence Theory

In Chapter 2, we showed that, at least in theory, the local linearization matrices are not derivatives. The formulas derived in Chapter 3 for the Lyapunov exponents are meaningful only if the numerically determined linearization matrices are close to the Eckmann-Ruelle linearizations $M(P)$. Because $M(P)$ is guaranteed to be the best linearization *in the limit* as the neighborhood radius shrinks to zero (see Theorems 2.2 and 2.6), it is possible that for a particular radius, a different matrix will do "better" than the Eckmann-Ruelle linearization, though we hope the "better" matrix will still be close to $M(P)$. In this chapter, we prove that, over small neighborhoods around the base point $P$, the best linearization matrix will in fact be close to $M(P)$.

To proceed, we must specify what we mean by the "best" linearization over a small neighborhood. That is, we need a means of measuring the error involved in the linearization process. The following definition serves this purpose by finding the worst-case error committed by a matrix $L$ used as the local linearization.

**Definition 4.1.** *Given $F : \mathbb{R}^m \to \mathbb{R}^m$, an trajectory of $F$, $P_0, P_1, \cdots \in \mathbb{R}^m$, and $P \in \mathbb{R}^m$ in the closure of the trajectory. For an $m \times m$ matrix $L$ and $\epsilon > 0$, define*

$$W(L, P, \epsilon) := \sup_{\|P_i - P\| \leq \epsilon} \|P_{i+1} - F(P) - L(P_i - P)\|$$

*where the supremum is taken over those values of $i$ for which $\|P_i - P\| \leq \epsilon$. For matrices $L_1$ and $L_2$, we say that $L_1$ is a* **better linearization** *than $L_2$ over the $\epsilon$-ball about $P$ provided $W(L_1, P, \epsilon) < W(L_2, P, \epsilon)$.*

In practice, one needs to avoid false-nearest-neighbors by taking the supremum over $i$ where both $\|P_i - P\| \leq \epsilon$ and $\|P_{i+1} - F(P)\| \leq \epsilon$. However, this is not necessary for the theory because $\pi$ is one-to-one in a neighborhood of $p$ (by either (A3') or (B3')) and we will let $\epsilon \to 0$. Hence, we can assume $\epsilon$ is small enough that the false-nearest-neighbor problem never arises.

We point out that the matrix minimizing $W(\cdot, P, \epsilon)$ is not necessarily the best least squares fit but rather the best "minimax" fit. The least squares problem

seems harder to formulate theoretically. At the end of this chapter, we present preliminary results for a least squares theory that incorporates infinite data.

In the statements of Theorems 4.2 and 4.8 below, we assume that we have a sequence of "best" linearizations $M_k$ over a shrinking set of $\epsilon_k$-balls. This is the interpretation we give to the condition $W(M_k, P, \epsilon_k) \leq W(M(P), P, \epsilon_k)$. Thus, each matrix $M_k$ does better than $M(P)$ over the $\epsilon_k$-ball around $P$ though not necessarily over any other ball around $P$.

As before, we begin with the one-dimensional case where $f : [0,1] \to [0,1]$ and $\pi : [0,1] \to \mathbb{R}^m$. We say the **reconstructed trajectory is dense in the curve** at the point $P \in \pi([0,1])$ provided there is a neighborhood $U$ of $P$ such that the trajectory $\{P_i : i \geq 0\}$ is dense in the curve $\pi([0,1]) \cap U$.

**Theorem 4.2 (Convergence Theorem, $\mathbb{R}^1 \to \mathbb{R}^m$).** *Assume (A1) and (A2). Let $P \in \pi([0,1])$ be a point satisfying (A3) and (A4) where the reconstructed trajectory is dense in the curve. Let $\{\epsilon_k\}_{k=1}^{\infty}$ be a decreasing sequence of positive numbers, $\epsilon_k \to 0$. Let $\{M_k\}_{k=1}^{\infty}$ be a sequence of matrices for which we have $W(M_k, P, \epsilon_k) \leq W(M(P), P, \epsilon_k)$ for all $k$. Then $M_k \to M(P)$ as $k \to \infty$.*

The key point for the proof is to control the size of $\|M_k - M(P)\|$ as $\epsilon_k \to 0$. This control will be achieved via the next proposition. Recall that a convex set contains every line segment connecting any two of its points. That is, $S$ is **convex** provided that whenever $u, v \in S$ and $t \in [0,1]$ we have $tu + (1-t)v \in S$. The **convex hull** of a set $S$ is the smallest convex set containing $S$.

**Proposition 4.3.** *Let $S$ be a subset of $\mathbb{R}^m$, and let $Hull(S)$ denote its convex hull. Assume that $Hull(S)$ contains a closed $m$-dimensional ball of radius $r$. If $A$ is an $m \times m$ matrix such that $\|Av\| \leq B$ for all $v \in S$, then $\|A\| \leq \frac{2B}{r}$.*

**Proof.** Note that $\|Av\| \leq B$ actually holds for all $v \in Hull(S)$. Let $\eta$ be a unit vector such that $\|A\eta\| = \|A\|$. Since $Hull(S)$ contains the closed ball $\overline{B(c,r)}$ for some $c \in Hull(S)$, there exists $z \in Hull(S)$ on the surface of this ball for which the vector $z - c$ points in the direction of $\eta$. Therefore, $z - c = r\eta$, and we have

$$r\,\|A\| = r\,\|A\eta\| = \|A(z-c)\| \leq \|Az\| + \|Ac\| \leq 2B.$$

□

Here is how we will use the proposition. We let $A_k$ be the matrix $M_k - M(P)$ and $S_k$ the set of small tangent vectors from the $\epsilon_k$-ball about $P$. The bound on $\|Av\|$ will come from the condition $W(M_k, P, \epsilon_k) \leq W(M(P), P, \epsilon_k)$. We will then be able to conclude a bound on $\|M_k - M(P)\|$ in terms of $\epsilon_k$. In order to proceed, then, we must examine the convex hull of the tangent vectors.

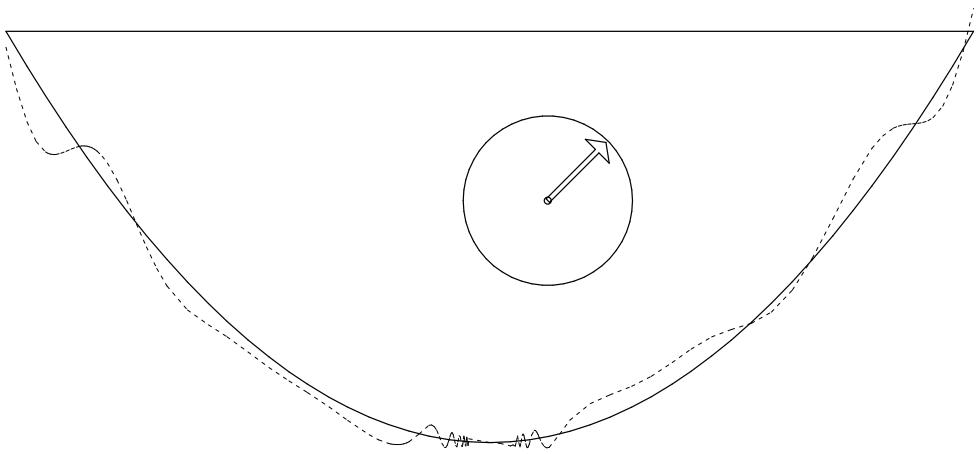


Figure 4.1: The solid curve outlines the convex hull of the curve $(h, \frac{1}{2}h^2)$ for $h \in [-1, 1]$. No small perturbation $V(h)$ (dashed curve) can significantly change the interior of this convex hull. Thus, the convex hull of $V(h)$ will always contain a ball of radius $C_{rad}$ .

Recall that when written in the canonical embedding basis at the base point $P$, little tangent vectors $\Delta P_i := P_i - P$ have the form shown in (2.2). This suggests that we consider curves in $\mathbb{R}^m$ of the form

$$V(h) = \left(h, \frac{1}{2}h^2, \ldots, \frac{1}{m!}h^m\right) + Pert(h) \tag{4.1}$$

where $Pert(h)$ is a continuous vector-valued function with $Pert(0) = 0$. Let $Hull_\epsilon(V)$ denote the convex hull of the set $\{V(h) : h \in [-\epsilon, \epsilon]\}$.

We need some additional notation. For $i = 0, 1, \ldots, m$, define $x_i = \frac{i}{m}$. Let $Vand_m$ denote the $m \times m$ Vandermonde matrix:

$$Vand_m = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \ldots & x_m^{m-1} \end{pmatrix}.$$

Since determinants vary continuously with the elements of the matrix, there is $\epsilon_0 > 0$ such that if we perturb the elements of $Vand_m$ by no more than $\epsilon_0$ then the determinant of the perturbed matrix, $Vand_m + E$, is still close to the determinant of the unperturbed matrix $Vand_m$:

$$\text{if } |E_{ij}| \le \epsilon_0 \text{ for all } i \text{ and } j, \text{ then } \det\left(Vand_m + E\right) \ge \frac{1}{2}\det\left(Vand_m\right) > 0.$$

**Lemma 4.4.** *There exists a constant $C_{rad} > 0$ such that for any curve $V(h)$ of the form (4.1) satisfying $\max_{-1\le h\le 1} \|Pert(h)\| \le \frac{\epsilon_0}{m\ m!}$, the convex hull $Hull_1(V)$ contains a ball of radius $C_{rad}$.*

**Proof.** The $m+1$ points $V(x_i)$ in $\mathbb{R}^m$, $i = 0, 1, \ldots, m$, form an $m$-simplex entirely contained (along with its interior) inside $Hull_1(V)$. Note that $V(x_0)$ is the zero vector. Treating the $V(x_i)$ as row-vectors and writing $\xi_{ij}$ for the $j$th component of $Pert(x_i)$, the volume of this simplex is given by [13, pp. 44]:

$$\begin{aligned}
\frac{1}{m!}\det\begin{vmatrix} V(x_1) \\ V(x_2) \\ \vdots \\ V(x_m) \end{vmatrix} &= \frac{1}{m!}\det\begin{vmatrix} x_1+\xi_{11} & \frac{1}{2}x_1^2+\xi_{12} & \ldots & \frac{1}{m!}x_1^m+\xi_{1m} \\ x_2+\xi_{21} & \frac{1}{2}x_2^2+\xi_{22} & \ldots & \frac{1}{m!}x_2^m+\xi_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_m+\xi_{m1} & \frac{1}{2}x_m^2+\xi_{m2} & \ldots & \frac{1}{m!}x_m^m+\xi_{mm} \end{vmatrix} \\
&= \frac{1}{m!}\frac{m!}{m^m}\left(\prod_{k=1}^{m}\frac{1}{k!}\right)\det\begin{vmatrix} 1+\frac{1}{x_1}\xi_{11} & x_1+\frac{2}{x_1}\xi_{12} & \ldots & x_1^{m-1}+\frac{m!}{x_1}\xi_{1m} \\ 1+\frac{1}{x_2}\xi_{21} & x_2+\frac{2}{x_2}\xi_{22} & \ldots & x_2^{m-1}+\frac{m!}{x_2}\xi_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 1+\frac{1}{x_m}\xi_{m1} & x_m+\frac{2}{x_m}\xi_{m2} & \ldots & x_m^{m-1}+\frac{m!}{x_m}\xi_{mm} \end{vmatrix} \\
&\ge \frac{1}{m^m}\left(\prod_{k=1}^{m}\frac{1}{k!}\right)\frac{1}{2}\det(Vand_m) > 0
\end{aligned}$$

because the perturbations of the elements of $Vand_m$ satisfy:

$$\left|\frac{j!}{x_i}\xi_{ij}\right| \le \frac{m!}{x_1}|\xi_{ij}| \le m\ m!\|Pert(x_i)\| \le \epsilon_0.$$

Since the vectors $V(x_i)$ cannot vary too far from their "unperturbed positions" at the points $\left(x_i, \frac{1}{2}x_i^2, \ldots, \frac{1}{m!}x_i^m\right)$, and since the $m$-simplex they form has an absolute lower bound on its volume, there must be some radius $C_{rad} > 0$ such that this $m$-simplex always contains a ball of that radius, no matter what the perturbation function $Pert(h)$. □

Note that the vectors $V(x_i)$ and the Vandermonde matrix used in this proof produce a simplex with very small volume. Indeed, other choices for the $x_i$ could produce larger volumes and hence larger "guaranteed" radii $C_{rad}$. Fortunately, we do not need the radius to be large, only that such a guaranteed radius does in fact exist.

**Proposition 4.5.** *Let $C_{Pert} > 0$ and $0 < \epsilon \le \min\left(1, \frac{\epsilon_0}{m\ m!C_{Pert}}\right)$. For any curve $V(h)$ of the form (4.1) with $\|Pert(h)\| \le C_{Pert}|h|^{m+1}$ defined on $[-\epsilon, \epsilon]$, $Hull_\epsilon(V)$ contains a ball of radius $C_{rad}\epsilon^m$, where $C_{rad}$ is independent of $\epsilon$ and $V(h)$.*

**Sketch of Proof.** This is a consequence of changing coordinate systems between the standard basis $\{e_1, e_2, \ldots, e_m\}$ and the basis $\{\epsilon e_1, \epsilon^2 e_2, \ldots, \epsilon^m e_m\}$. Then, as $h$ ranges over $[-\epsilon, \epsilon]$, $\frac{h}{\epsilon}$ ranges over $[-1, 1]$. Together, the conditions on $\epsilon$ and $Pert(h)$ imply that the perturbation term is still small enough to use Lemma 4.4. This guarantees the existence of a ball of radius $C_{rad}$ contained within the convex hull of $V\left(\frac{h}{\epsilon}\right)$ over $[-1, 1]$. Naturally, this ball has axes of length $C_{rad}$ in each of the principal directions of the second basis. When we convert coordinates back to the standard basis, the ball becomes an ellipsoid with axes of length $C_{rad}\epsilon^i$ for $i = 1, 2, \ldots, m$. In particular, the ellipsoid contains a ball of radius $C_{rad}\epsilon^m$. □

**Proof of Theorem 4.2.** The hypotheses guarantee that the Eckmann-Ruelle linearization $M(P)$ exists and is the unique best linearization as $\epsilon_k \to 0$. We will show that there is a constant $C$ (independent of $k$) such that $\|M_k - M(P)\| \leq C\epsilon_k$ for large $k$.

By Theorem 2.2, there are $\epsilon' > 0$ and $C_{er} > 0$ such that for every $0 < \epsilon < \epsilon'$ we have $W(M(P), P, \epsilon) \leq C_{er}\epsilon^{m+1}$. Without loss of generality, $\epsilon_1 < \min\{\epsilon', 1\}$. Then,

$$W(M_k, P, \epsilon_k) \leq W(M(P), P, \epsilon_k) \leq C_{er}\epsilon_k^{m+1} \quad \text{for } k = 1, 2, \ldots \tag{4.2}$$

First, we construct a suitable set $S_k$. With respect to the canonical embedding basis at $P$, $\Delta P_i := P_i - P = \left(h_i, \frac{1}{2}h_i^2, \ldots, \frac{1}{m!}h_i^m\right)_P + Rem(h_i)$. Without loss of generality we may assume that $Rem(h)$ is defined on $[-\epsilon_1, \epsilon_1]$. Define the function $V(h)$ on $[-\epsilon_1, \epsilon_1]$ by equation (4.1) with $Pert(h_i) = Rem(h_i)$, and note that $\Delta P_i = V(h_i)$. Since the reconstructed trajectory is dense in the curve at $P$, we may assume that the points $P_i$ in $B(P, \epsilon_1)$ trace out the curve $\pi([0, 1]) \cap B(P, \epsilon_1)$. Because $\pi$ is a diffeomorphism, these points $P_i \in I\!R^m$ pull back to scalars $p_i \in I\!R$ that are dense in some subinterval of $[0, 1]$. By (2.1), this interval contains the subinterval $\left[p - \frac{\epsilon_1}{C_\pi}, p + \frac{\epsilon_1}{C_\pi}\right]$. It follows that the $h_i = p_i - p$ are dense in the interval $I_1 = \left[-\frac{\epsilon_1}{C_\pi}, \frac{\epsilon_1}{C_\pi}\right]$. By similar reasoning, for each $k$, the preimages of $P_i \in B(P, \epsilon_k) \cap \pi([0, 1])$ are dense in the interval $\left[p - \frac{\epsilon_k}{C_\pi}, p + \frac{\epsilon_k}{C_\pi}\right]$, and the corresponding $h_i$ are dense in $I_k = \left[-\frac{\epsilon_k}{C_\pi}, \frac{\epsilon_k}{C_\pi}\right]$. Define the set $S_k := \{V(h) : h \in I_k\}$ in $I\!R^m$. The convex hull of $S_k$ is $Hull_{\frac{\epsilon_k}{C_\pi}}(V)$, and by Proposition 4.5, it contains a ball of radius $\frac{C_{rad}}{C_\pi^m}\epsilon_k^m$ whenever $\epsilon_k \leq \min\left(1, \frac{\epsilon_0}{m\ m!C_{Taylor}}\right)$.

For each $k$, let $A_k := M_k - M(P)$ be the $m \times m$ error matrix. Then, for each $i$ with $\|\Delta P_i\| \leq \epsilon_k$, we have by (4.2)

$$\begin{aligned}
\|A_k \Delta P_i\| &\leq \|P_{i+1} - F(P) - M(P)\Delta P_i\| + \|P_{i+1} - F(P) - M_k \Delta P_i\| \\
&\leq W(M(P), P, \epsilon_k) + W(M_k, P, \epsilon_k) \\
&\leq 2C_{er}\epsilon_k^{m+1}
\end{aligned}$$

Because $\Delta P_i = V(h_i)$, the $h_i$ are dense in $I_k$, and $V(h)$ is continuous on this interval, this inequality extends to all $h \in I_k$. Therefore,

$$\|A_k v\| \leq 2C_{er}\epsilon_k^{m+1} \qquad \text{for all } v \in S_k,$$

and, by Proposition 4.3,

$$\|A_k\| \leq \frac{4C_{er}C_\pi^m}{C_{rad}}\epsilon_k.$$

This completes the proof. □

Before we proved Theorem 4.2 in the one-dimensional case, we added an assumption about the distribution of the orbit within the reconstructed attractor that was not needed in previous chapters. In particular, we assumed that, in a neighborhood of the base point, the orbit was dense along the curved segment of the reconstructed attractor. Unfortunately, a density assumption is much too strong for the case of two-dimensional underlying dynamics because our reconstructed attractor could be fractal in nature.

It turns out that we need only add a third condition to the definition of "approach direction." This condition ensures that the distances between the base point and points approaching along the direction vector shrink exponentially fast.

**Definition 4.6.** *Let $l$ be a unit vector in $\mathbb{R}^m$. A subset $S$ of $\mathbb{R}^m$ has the* **fractal approach direction** *$l$ at the base point $P \in \mathbb{R}^m$ if there is a sequence $\{Q_k\}_{k=1}^{\infty}$ from $S$ such that:*

1. *$Q_k \to P$ as $k \to \infty$,*
2. *$\frac{\Delta Q_k}{\|\Delta Q_k\|} \to l$ as $k \to \infty$, where $\Delta Q_k := Q_k - P$, and*
3. *there are constants $0 < C^- \leq C^+ < 1$ such that for all $k$*

$$C^- \leq \frac{\|\Delta Q_{k+1}\|}{\|\Delta Q_k\|} \leq C^+. \tag{4.3}$$

*A collection of fractal approach directions at $P$ is* **distinct** *provided that no two are the same and no two are reflections through the origin. Multiple fractal approach directions are not required to be linearly independent.*

As noted in Chapter 2 (see Figure 2.2), this kind of behavior occurs for points in hyperbolic systems and may occur in all chaotic systems with two or more dimensions.

We need a lemma like 2.5 that translates fractal approach directions in the reconstruction space down to fractal approach directions in the underlying phase space.

**Lemma 4.7.** *Assume (B1) and (B2). Let $P = \pi(p)$ be a point of $\pi([0,1] \times [0,1])$ for which (B3) and (B4) hold. If $P$ is in the closure of a trajectory of $F$, $P_0, P_1, \cdots \in \mathbb{R}^5$ that has $d$ distinct fractal approach directions at $P$, then the underlying trajectory $p_0, p_1, \cdots \in \mathbb{R}^2$, where $P_i = \pi(p_i)$, has $d$ distinct fractal approach directions at the point $p \in [0,1] \times [0,1]$.*

**Proof.** By Lemma 2.5, we get $d$ distinct approach directions in $\mathbb{R}^2$. Thus, we need only verify the third condition for each one. Let $l$ be an approach direction in $\mathbb{R}^2$ coming from a fractal approach direction in $\mathbb{R}^5$. Let $Q_k = \pi(q_k)$, $k = 1, 2, \ldots$, be the points in $\mathbb{R}^5$ forming the subsequence that converges to this fractal approach direction. By the previous comment, $q_k \to p$ and $\frac{q_k - p}{\|q_k - p\|} \to l$. We can use the constants $C^-, C^+$ from condition (3) together with (2.1) to translate (3) to:

$$\frac{C^-}{C_\pi^2} \leq \frac{\|q_{k+1} - p\|}{\|q_k - p\|} \leq C^+ C_\pi^2.$$

If $C^+ C_\pi^2 < 1$, we are done. Otherwise, we have to come up with new bounds. We form a new subsequence from $\{q_k\}_{k=1}^\infty$ as follows. Start with $\hat{q}_1 = q_1$. Then, having picked $\hat{q}_k = q_i$ for some $i$, we must pick $\hat{q}_{k+1}$. Look back to the original sequence $\{q_k\}$ and set $\hat{q}_{k+1}$ to be the first term after $q_i$, say $q_{i+j}$ with $j > 0$, for which $\|q_{i+j} - p\| < C^+ \|q_i - p\|$. Note that $\|q_{i+j-1} - p\| \geq C^+ \|q_i - p\|$. Then,

$$C^+ > \frac{\|\hat{q}_{k+1} - p\|}{\|\hat{q}_k - p\|} = \frac{\|q_{i+j} - p\|}{\|q_{i+j-1} - p\|} \frac{\|q_{i+j-1} - p\|}{\|q_i - p\|} \geq \frac{C^-}{C_\pi^2} C^+.$$

These bounds now satisfy (3) of the definition. Of course, since $\{\hat{q}_k\}$ is a subsequence of $\{q_k\}$, we have $\hat{q}_k \to p$ and $\frac{\hat{q}_k - p}{\|\hat{q}_k - p\|} \to l$ as well. □

The Convergence Theorem below applies only for the case of a reconstruction in $\mathbb{R}^5$, but the extensions to appropriate higher-dimensional cases should be clear.

**Theorem 4.8 (Convergence Theorem, $\mathbb{R}^2 \to \mathbb{R}^5$).** *Assume (B1) and (B2). Let $P = \pi(p)$ be a point of $\pi([0,1] \times [0,1])$ for which (B3) and (B4) hold. Assume that $P$ is in the closure of a trajectory of $F$, $P_0, P_1, \cdots \in \mathbb{R}^5$ that has three distinct fractal approach directions at $P$. Let $\{\epsilon_k\}_{k=1}^\infty$ be a decreasing sequence of positive numbers, $\epsilon_k \to 0$. Let $\{M_k\}_{k=1}^\infty$ be a sequence of matrices such that $W(M_k, P, \epsilon_k) \leq W(M(P), P, \epsilon_k)$ for all $k$. Then $M_k \to M(P)$ as $k \to \infty$.*

This proof will be similar to that given in the one-dimensional case over the course of Lemmas 4.3, 4.4, and 4.5. The present situation will be far more technically involved, however, because the geometry is no longer simple. In the previous case, we had only a single approach direction and it was reasonable to assume the trajectory was dense along the line. These properties are not generic when

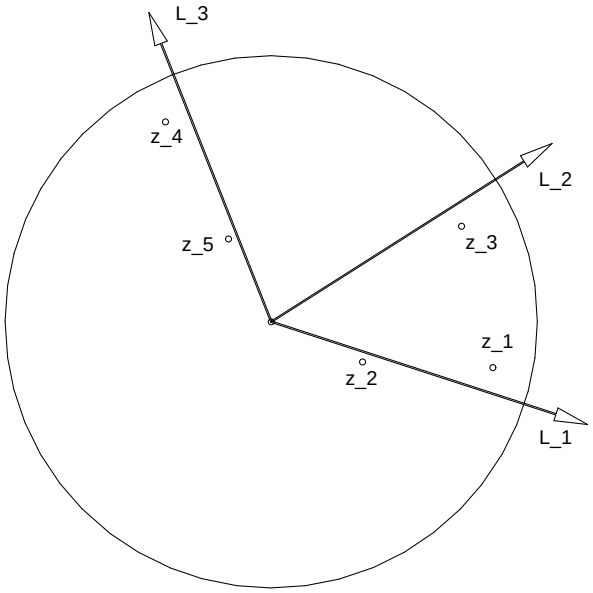



Figure 4.2: For this configuration of points $z_1, \ldots, z_5$ in $\mathbb{R}^2$, the convex hull in $\mathbb{R}^5$ of the set $\{\pi(z_1), \ldots, \pi(z_5)\}$ is guaranteed to contain a ball of radius $C_{rad}$.

the underlying dynamics have two or more dimensions. (See Figure 2.2.) In the present case, the local distribution of the trajectory can be sparse, approaching only along three fractal approach directions. Recall from the discussion in Chapter 2 that three approach directions are necessary to ensure that we can uniquely determine the Eckmann-Ruelle linearization.

In Lemmas 4.4 and 4.5, we singled out a configuration of points in the underlying space (easily chosen thanks to the density assumption) whose reconstructed images had a convex hull that could be guaranteed to contain a small ball. We will do the same thing here, though we have to be more careful. Our points are no longer nicely constrained to a line (they converge along three approach directions in the plane), and we cannot choose any points we want (there is no density assumption). We will choose five points $z_i$ situated near our three fractal approach directions $l_i$ at $P$ as in Figure 4.2. We will use the spacing between points $z_i$ that (4.3) guarantees in the same way that we used the guaranteed spacing between $x_i = \frac{i}{m}$ and $x_{i-1} = \frac{i-1}{m}$ in the one-dimensional case.

To be more specific, for $i = 1, 2, 3$, let $l_i = (\cos(\alpha_i), \sin(\alpha_i))$ be the fractal approach directions. Without loss of generality, we can label the $l_i$ to satisfy

$$\begin{aligned} 0 \leq \alpha_i < 2\pi \qquad & \text{for } i = 1, 2, 3 \\ 0 < \alpha_2 - \alpha_1 < \pi & \\ 0 < \alpha_3 - \alpha_2 < \pi & \end{aligned} \tag{4.4}$$

Each $l_i$ will have associated constants $C_i^-$ and $C_i^+$ from (4.3), but we can easily

find constants $0 < C^- \le C^+ < 1$ that work for all three $l_i$ simultaneously. We want to consider quintuples of points $z_i = (r_i \cos(\theta_i), r_i \sin(\theta_i)) \in \mathbb{R}^2$, $1 \le i \le 5$, that satisfy these constraints for some $\eta > 0$:

$$\begin{array}{ll} C^- \le r_i \le 1 & i = 1, 3, 4 \\ C^- \le \frac{r_{i+1}}{r_i} \le C^+ & i = 1, 4 \end{array} \qquad \begin{array}{ll} |\theta_i - \alpha_1| \le \eta & i = 1, 2 \\ |\theta_3 - \alpha_2| \le \eta & \\ |\theta_i - \alpha_3| \le \eta & i = 4, 5 \end{array} \tag{4.5}$$

We will also consider functions $V : \mathbb{R}^2 \to \mathbb{R}^5$ of the form

$$V(x, y) = \left(x, y, \frac{1}{2}x^2, xy, \frac{1}{2}y^2\right) + Pert(x, y) \tag{4.6}$$

where $Pert(x, y)$ is a continuous vector-valued function. In the following lemma, the function $V$ plays the role of our measurement function $\pi$, written in canonical embedding coordinates.

**Lemma 4.9.** *Given constants $0 < C^- \le C^+ < 1$ and $\alpha_1, \alpha_2, \alpha_3$ satisfying (4.4), there exist positive constants $\eta_0$, $\rho$, and $C_{rad}$ such that for*

1. *any five points $z_i \in \mathbb{R}^2$, $1 \le i \le 5$, satisfying (4.5) with $0 < \eta \le \eta_0$, and*
2. *any function $V : \mathbb{R}^2 \to \mathbb{R}^5$ of the form (4.6) with $Pert(0, 0) = 0$ and $\|Pert(x, y)\| \le \rho$ for all $\|(x, y)\| \le 1$,*

*the convex hull in $\mathbb{R}^5$ of $\{V(0), V(z_1), \ldots, V(z_5)\}$ contains a ball of radius $C_{rad}$.*

**Proof.** Given the five points $z_i \in \mathbb{R}^2$ satisfying (4.5), we define the $5 \times 5$ matrix $N = N(r_1, \ldots, r_5, \theta_1, \ldots, \theta_5)$:

$$N = \begin{pmatrix} r_1\cos(\theta_1) & r_1\sin(\theta_1) & \frac{1}{2}r_1^2\cos^2(\theta_1) & r_1^2\cos(\theta_1)\sin(\theta_1) & \frac{1}{2}r_1^2\sin^2(\theta_1) \\ r_2\cos(\theta_2) & r_2\sin(\theta_2) & \frac{1}{2}r_2^2\cos^2(\theta_2) & r_2^2\cos(\theta_2)\sin(\theta_2) & \frac{1}{2}r_2^2\sin^2(\theta_2) \\ r_3\cos(\theta_3) & r_3\sin(\theta_3) & \frac{1}{2}r_3^2\cos^2(\theta_3) & r_3^2\cos(\theta_3)\sin(\theta_3) & \frac{1}{2}r_3^2\sin^2(\theta_3) \\ r_4\cos(\theta_4) & r_4\sin(\theta_4) & \frac{1}{2}r_4^2\cos^2(\theta_4) & r_4^2\cos(\theta_4)\sin(\theta_4) & \frac{1}{2}r_4^2\sin^2(\theta_4) \\ r_5\cos(\theta_5) & r_5\sin(\theta_5) & \frac{1}{2}r_5^2\cos^2(\theta_5) & r_5^2\cos(\theta_5)\sin(\theta_5) & \frac{1}{2}r_5^2\sin^2(\theta_5) \end{pmatrix}$$

and a function $\mu$:

$$\mu(r_1, \ldots, r_5, \theta_1, \ldots, \theta_5) = \frac{1}{5!} \det N(r_1, \ldots, r_5, \theta_1, \ldots, \theta_5).$$

Note that the constraints in (4.5) imply

$$\begin{aligned} r_2 &\ge C^- r_1 r_1 - r_2 \ge (1 - C^+) r_1 \\ r_5 &\ge C^- r_4 r_4 - r_5 \ge (1 - C^+) r_4 \end{aligned}$$

so that

$$\begin{aligned}
&\mu(r_1, r_2, r_3, r_4, r_5, \alpha_1, \alpha_1, \alpha_2, \alpha_3, \alpha_3) \\
&= \frac{1}{4 \cdot (5!)} r_1 r_2 r_3^2 r_4 r_5 (r_1 - r_2)(r_4 - r_5) \sin(\alpha_2 - \alpha_1) \sin^2(\alpha_3 - \alpha_1) \sin(\alpha_3 - \alpha_2) \\
&\geq \frac{1}{480} \left(C^-\right)^{10} (1 - C^+)^2 \sin(\alpha_2 - \alpha_1) \sin^2(\alpha_3 - \alpha_1) \sin(\alpha_3 - \alpha_2) \\
&:= B(C^-, C^+, \alpha_1, \alpha_2, \alpha_3)
\end{aligned}$$

Since determinants vary continuously with the matrix elements, there exists $\eta_0$ such that if the $z_i$ satisfy (4.5) with $0 < \eta \leq \eta_0$, we have:

$$\begin{aligned}
\mu(r_1, r_2, r_3, r_4, r_5, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5) &\geq \frac{1}{2} \mu(r_1, r_2, r_3, r_4, r_5, \alpha_1, \alpha_1, \alpha_2, \alpha_3, \alpha_3) \\
&\geq \frac{1}{2} B(C^-, C^+, \alpha_1, \alpha_2, \alpha_3)
\end{aligned}$$

In particular, when $0 < \eta \leq \eta_0$, we can bound $\mu$ from below independent of $r_i$, $\theta_i$, and $\eta$. It follows that there exists $\rho > 0$ such that if we perturb any matrix $N = N(r_1, \ldots, r_5, \theta_1, \ldots, \theta_5)$ by an error matrix $E = (E_{ij})$ with $|E_{ij}| \leq \rho$, then $\det(N + E) \geq \frac{1}{2} \det(N)$.

Since $V(0, 0) = 0$, the simplex formed in $\mathbb{R}^5$ from the six points $V(0)$, $V(z_1)$, $\ldots$, $V(z_5)$ has five-dimensional volume given by the determinant [13, pp. 44]:

$$\frac{1}{5!} \det \begin{pmatrix} V(z_1) \\ \vdots \\ V(z_5) \end{pmatrix} \geq \frac{1}{2} \frac{1}{5!} \det \left( N(r_1, \ldots, r_5, \theta_1, \ldots, \theta_5) \right) \geq \frac{1}{4} B(C^-, C^+, \alpha_1, \alpha_2, \alpha_3)$$

whenever the $z_i$ satisfy (4.5) with $0 < \eta \leq \eta_0$ and $\|Pert(x, y)\| \leq \rho$. Thus, the convex hull of these points has a guaranteed minimum volume. Since the positions of the points $V(z_i)$ are constrained in space, there is a constant $C_{rad} > 0$ such that there will always be a small ball of radius $C_{rad}$ contained somewhere within this convex hull. □

**Lemma 4.10.** *Let $C^-$, $C^+$, $\alpha_1$, $\alpha_2$, $\alpha_3$, $C_{rad}$, $\eta_0$, and $\rho$ be as in Lemma 4.9. Let $C_{Pert} > 0$ and $0 < \epsilon \leq \min\left(1, \frac{\rho}{C_{Pert}}\right)$. Then, for*

1. *any five points $z_i = (\epsilon r_i \cos(\theta_i), \epsilon r_i \sin(\theta_i))$ in $\mathbb{R}^2$, $1 \leq i \leq 5$, with $r_i$ and $\theta_i$ satisfying (4.5) for some $0 < \eta \leq \eta_0$, and*

2. *any function $V : \mathbb{R}^2 \to \mathbb{R}^5$ of the form (4.6) with*

$$\|Pert(x, y)\| \leq C_{Pert} \|(x, y)\|^3 \qquad \text{for all } \|(x, y)\| \leq \epsilon,$$

*the convex hull in $\mathbb{R}^5$ of $\{V(0), V(z_1), \ldots, V(z_5)\}$ contains a ball of radius $C_{rad}\epsilon^2$.*

**Sketch of Proof.** This proof is similar to that of Lemma 4.5 and follows from changing coordinate systems between the standard basis $\{e_1, e_2, e_3, e_4, e_5\}$ and the basis $\{\epsilon e_1, \epsilon e_2, \epsilon^2 e_3, \epsilon^2 e_4, \epsilon^2 e_5\}$. □

**Proof of Theorem 4.8.** The hypotheses guarantee that the Eckmann-Ruelle linearization $M(P)$ exists and is the unique best linearization as $\epsilon_k \to 0$. We will show that there is a constant $C$ (independent of $k$) such that $\|M_k - M(P)\| \leq C\epsilon_k$ for large $k$.

By Theorem 2.6, there are $\epsilon' > 0$ and $C_{er} > 0$ such that for every $0 < \epsilon < \epsilon'$ we have $W(M(P), P, \epsilon) \leq C_{er}\epsilon^3$. Without loss of generality, $\epsilon_1 < \min\{\epsilon', 1\}$. Then,

$$W(M_k, P, \epsilon_k) \leq W(M(P), P, \epsilon_k) \leq C_{er}\epsilon_k^3 \quad \text{for } k = 1, 2, \ldots$$

For each $k$, let $A_k := M_k - M(P)$ and let $S_k$ be the set of small tangent vectors at $P$:

$$S_k := \left\{ P_j - P \in \mathbb{R}^5 : P_j = \pi(p_j) \text{ and } \|p_j - p\| \leq \frac{\epsilon_k}{C_\pi} \right\}.$$

For each $P_j \in S_k$, $\|P_j - P\| \leq \epsilon_k$ by (2.1). As in the proof of Theorem 4.2, we see that

$$\|A_k v\| \leq 2C_{er}\epsilon_k^3 \qquad \text{for each } v \in S_k.$$

Thus, we need only show that the convex hull $Hull(S_k)$ contains a ball in $\mathbb{R}^5$. We will use Lemma 4.10 to accomplish this.

With respect to the canonical embedding basis at $P$, we have for $P_i$ near $P$:

$$P_i - P = \left( h_{i1}, h_{i2}, \frac{1}{2}h_{i1}^2, h_{i1}h_{i2}, \frac{1}{2}h_{i2}^2 \right)_P + Rem(h_{i1}, h_{i2}, p)$$

where $p_i - p = (h_{i1}, h_{i2}) \in \mathbb{R}^2$ and $Rem(h_{i1}, h_{i2}, p)$ is a vector-valued function continuous in a neighborhood containing the closed ball of radius $\frac{\epsilon_1}{C_\pi}$ about (0,0). Define the function $V(h_1, h_2)$ as in Lemma 4.10, with $Pert(h_1, h_2) = Rem(h_1, h_2, p)$. Note that $P_i - P = V(h_{i1}, h_{i2})$ and we can take $C_{Pert} = C_{Taylor}$ since $\|Rem(h_{i1}, h_{i2}, p)\| \leq C_{Taylor} \|(h_{i1}, h_{i2})\|^3$ (see Chapter 2). It remains to construct the five points in $\mathbb{R}^2$ that Lemma 4.10 requires.

By hypothesis, the attractor has three distinct fractal approach directions, and by Lemma 4.7, these translate down to distinct fractal approach directions $l_1, l_2, l_3$ at $p \in \mathbb{R}^2$ in the two-dimensional underlying attractor. These approach directions in $\mathbb{R}^2$ are unit vectors, so we write $l_i = (\cos(\alpha_i), \sin(\alpha_i))$ for $i = 1, 2, 3$. Since they are distinct approach directions, the angles $\alpha_i$ are distinct and no two differ by a multiple of $\pi$. Moreover, we can arrange the $l_i$ so that they satisfy (4.4). For each $l_i$, there is a sequence of trajectory points in $\mathbb{R}^2$, $q_{ij} = p_{n(i,j)}$, $j = 1, 2, \ldots,$

satisfying the definition of fractal approach direction for that $l_i$. Thus, there are constants $0 < C^- \leq C^+ < 1$ such that

$$C^- \leq \frac{\left\| q_{i(j+1)} - p \right\|}{\|q_{ij} - p\|} \leq C^+ \qquad \text{for each } i = 1, 2, 3 \text{ and } j = 1, 2, \ldots \tag{4.7}$$

where the constants $C^-$ and $C^+$ work for all three fractal approach directions. At this point, we have constants $C^-$, $C^+$, $\alpha_1$, $\alpha_2$, and $\alpha_3$, and we obtain the constants $C_{rad}$, $\eta_0$ and $\rho$ from Lemma 4.9.

It will be convenient to write the points $q_{ij}$ in coordinates with origin at $p$: $q_{ij} - p = (r_{ij}\cos(\theta_{ij}), r_{ij}\sin(\theta_{ij}))$. From condition (2) of the definition of fractal approach direction, we have $\frac{q_{ij}-p}{\|q_{ij}-p\|} \to l_i$ as $j \to \infty$, which translates naturally to $\theta_{ij} \to \alpha_i$. We may now choose a number $J$ large enough that for any $i = 1, 2, 3$ and $j \geq J$, $|\theta_{ij} - \alpha_i| \leq \eta_0$. Let $K$ denote the least index $k$ for which $\frac{\epsilon_k}{C_\pi} \leq \min\left(1, \frac{\rho}{C_{Pert}}\right)$ and such that for each $i = 1, 2, 3$, we have $\max\{\|q_{ij} - p\| : j \geq J\} > \frac{\epsilon_k}{C_\pi}$.

Now, let $k \geq K$ and consider the $\frac{\epsilon_k}{C_\pi}$-ball about $p$. Let $q_{1j}$, $j \geq J$, be the first point in the sequence converging to $l_1$ for which $\|q_{1j} - p\| \leq \frac{\epsilon_k}{C_\pi}$. Set $z_1 = q_{1j} - p = (r_1\cos(\theta_1), r_1\sin(\theta_1))$. Since $\|q_{1(j-1)} - p\| > \frac{\epsilon_k}{C_\pi}$, it follows from (4.7) that $r_1 = \|q_{1j} - p\| \geq \frac{\epsilon_k}{C_\pi} C^-$. Let $z_2 = q_{1(j+1)} - p = (r_2\cos(\theta_2), r_2\sin(\theta_2))$ correspond to the next point in the sequence for $l_1$. Thus, $C^- \leq \frac{r_2}{r_1} \leq C^+$. Similarly, we choose points $z_4$ and $z_5$ for $l_3$. We choose $z_3$ from $l_2$ in the same way we chose $z_1$ and $z_4$, i.e., to be the first point along the subsequence converging to $l_2$ within the radius of $\frac{\epsilon_k}{C_\pi}$. We do not need a closer point along $l_2$. Figure 4.2 shows the configuration we have constructed.

Let $\hat{r}_i = \frac{C_\pi}{\epsilon_k} r_i$ and write $z_i = \left(\frac{\epsilon_k}{C_\pi}\hat{r}_i\cos(\theta_i), \frac{\epsilon_k}{C_\pi}\hat{r}_i\sin(\theta_i)\right)$ for $i = 1, \ldots, 5$. By construction, the $\hat{r}_i$ and $\theta_i$ satisfy (4.5) with $\eta = \eta_0$. Thus, we can apply Lemma 4.10 to conclude that the convex hull of the points $V(0) = P$, $V(z_1)$, $\ldots$ , $V(z_5)$ contains a five-dimensional ball of radius $\frac{C_{rad}}{C_\pi^2}\epsilon_k^2$. Of course, each point $V(z_i)$ corresponds to some vector $P_j - P$ in $S_k$, and it follows that $Hull(S_k)$ contains this small ball. Finally, by Proposition 4.3, we conclude that

$$\|A_k\| \leq \frac{4C_{er}C_\pi^2}{C_{rad}}\epsilon_k \qquad \text{for } k \geq K.$$

□

We turn our attention now to formulating our convergence theorems in terms of least squares estimates rather than minimax estimates. As before, we need a way of comparing matrices, i.e., a way of measuring the error involved in the linearization process over a particular $\epsilon$-ball. In practice, this is done by summing the squares of the error terms. The techniques of least squares can then determine

the matrix which minimizes this sum of squared errors. Usually the data set is finite. However, we are interested in a convergence question which, by its very nature, requires an infinite data set. Thus, we must find a reasonable notion for least squares that allows us to consider infinite data. The arguments we give now are intended to motivate the definition that will follow.

Imagine that we have exactly $N$ trajectory points $P_0, \ldots, P_{N-1}$ in $\mathbb{R}^m$ where $F(P_i) = P_{i+1}$, and we want to determine the best local linearization in the least squares sense over an $\epsilon$-neighborhood about $P$. We would then be looking for a matrix $M$ which minimizes

$$\sum_{\|P_i - P\| \leq \epsilon} \|P_{i+1} - F(P) - M(P_i - P)\|^2$$

where the sum is taken over only those trajectory points with $\|P_i - P\| \leq \epsilon$. Any matrix which minimizes the quantity above will also minimize

$$\frac{1}{N'} \sum_{i=0}^{N-1} \|P_{i+1} - F(P) - M(P_i - P)\|^2 \chi_{B(P,\epsilon)}(P_i)$$

where $N'$ is the number of data points $\epsilon$-close to $P$. This quantity represents the average of the squared errors. The characteristic function $\chi_{B(P,\epsilon)}$ is used to selectively pick off just those values which are close to the base point $P$. For the infinite data case, we take the limit as $N \to \infty$, being careful to adjust the value of $N'$ for each value of $N$:

$$\lim_{N\to\infty} \frac{1}{N'} \sum_{i=0}^{N-1} \|P_{i+1} - F(P) - M(P_i - P)\|^2 \chi_{B(P,\epsilon)}(P_i) \tag{4.8}$$

where

$$N' = \sum_{i=0}^{N-1} \chi_{B(P,\epsilon)}(P_i).$$

The number $N'$ counts those trajectory points $P_i$ (among the first $N$ iterates) which are $\epsilon$-close to $P$. We will rewrite (4.8) using the Birkhoff Ergodic Theorem [14]. In what follows, we let $A$ be the attractor of the underlying system and $\mu$ its natural measure (an invariant, ergodic probability measure). Set $I(P,\epsilon) = A \cap \pi^{-1}(B(P,\epsilon))$. For a matrix $L$, we define the differentiable function

$$\Phi_L(x,p) := \|\pi(f(x)) - \pi(f(p)) - L(\pi(x) - \pi(p))\|^2.$$

For almost every $p_0$ (with respect to $\mu$), the limit (4.8) can be written:

$$\begin{aligned}\lim_{N\to\infty} \frac{\frac{1}{N}\sum_{i=0}^{N-1} \Phi_L(f^i(p_0),p)\chi_{B(P,\epsilon)}(\pi(f^i(p_0)))}{\frac{1}{N}\sum_{i=0}^{N-1} \chi_{B(P,\epsilon)}(\pi(f^i(p_0)))} &= \frac{\int_A \Phi_L(x,p)\chi_{B(P,\epsilon)}(\pi(x))d\mu(x)}{\int_A \chi_{B(P,\epsilon)}(\pi(x))d\mu(x)} \\ &= \frac{1}{\mu(I(P,\epsilon))} \int_{I(P,\epsilon)} \Phi_L(x,p)d\mu(x)\end{aligned}$$

Since the denominator $\mu(I(P,\epsilon))$ is independent of the matrix $L$, any matrix which minimizes the limit in (4.8) also minimizes the quantity

$$LS(L,P,\epsilon) := \left(\int_{I(P,\epsilon)} \Phi_L(x,\pi^{-1}(P))d\mu(x)\right)^{\frac{1}{2}}.$$

We say the matrix $L_1$ is a **better linearization** of $F$ over the $\epsilon$-ball about $P$ than the matrix $L_2$ provided $LS(L_1,P,\epsilon) < LS(L_2,P,\epsilon)$.

For the Eckmann-Ruelle matrix $M(P)$ of a reconstruction from $\mathbb{R}^1$ into $\mathbb{R}^m$, it follows from the proof of the Local Linearization Theorem 2.2 that

$$\begin{aligned}\int_{I(P,\epsilon)} &\Phi_{M(P)}(x,\pi^{-1}(P))d\mu(x) \\ &= \int_{I(P,\epsilon)} \|Rem_f(x-p) - M(P)Rem(x-p)\|^2 \, d\mu(x) \\ &\le C_{Taylor}^2 \left(1+\|M(P)\|\right)^2 \int_{I(P,\epsilon)} |x-p|^{2m+2} d\mu(x) \\ &\le C_{Taylor}^2 \left(1+\|M(P)\|\right)^2 C_\pi^{2m+2}\epsilon^{2m+2}\mu\left(I(P,\epsilon)\right)\end{aligned}$$

since $|x-p| \le C_\pi \|\pi(x)-P\| \le C_\pi\epsilon$ by (2.1). Because the natural measure $\mu$ is a probability measure,

$$LS(M(P),P,\epsilon) \le C\epsilon^{m+1}. \tag{4.9}$$

We can do better still if $\mu$ is a uniform measure on the unit interval $[0,1]$, that is $\mu$ is absolutely continuous with respect to Lebesgue measure and has a positive continuous density function $\mathcal{F}(x)$. In this case, since $I(P,\epsilon)$ is contained in the subinterval $[p-C_\pi\epsilon, p+C_\pi\epsilon]$ by (2.1), we have

$$\mu(I(P,\epsilon)) \le \int_{p-C_\pi\epsilon}^{p+C_\pi\epsilon} d\mu(x) \le \max_{0\le x\le 1} |\mathcal{F}(x)| \cdot 2C_\pi\epsilon.$$

Thus, we can improve (4.9) in this case to

$$LS(M(P),P,\epsilon) \le C\epsilon^{m+\frac{3}{2}}. \tag{4.10}$$

The goal for the rest of this chapter is to prove that, in the case where $\mu$ is a uniform measure on [0,1], any sequence of matrices will converge to $M(P)$ if each matrix in the sequence is a better linearization than $M(P)$ over some small ball. Specifically, we prove:

**Theorem 4.11 (Least Squares Convergence, $\mathbb{R}^1 \to \mathbb{R}^m$).** *Assume (A1) and (A2). Let $A$ be a compact attractor for $f$ with natural measure $\mu$. Assume that $\mu$ is invariant, ergodic, and uniform, with positive density function $\mathcal{F}$. Let $P \in \mathbb{R}^m$ be a limit point of the reconstructed attractor $\pi(A)$ for which (A3) and (A4) hold. Let $\{\epsilon_k\}_{k=1}^\infty$ be a decreasing sequence of positive numbers, $\epsilon_k \to 0$. Let $\{M_k\}_{k=1}^\infty$ be a sequence of matrices such that $LS(M_k,P,\epsilon_k) \le LS(M(P),P,\epsilon_k)$ for all $k$. Then $M_k \to M(P)$ as $k \to \infty$.*

Before giving the proof of 4.11, we prove a few lemmas. In what follows, $\langle \cdot, \cdot \rangle$ will denote the usual inner product of vectors in $\mathbb{R}^m$. We start with a definition.

**Definition 4.12.** *For a function $f$ mapping the interval [-1,1] into $\mathbb{R}^m$, and for $0 \leq \epsilon \leq 1$, we define*

$$\delta(f, \epsilon) := \min_{\|\eta\|=1} \int_{-\epsilon}^{\epsilon} \langle \eta, f(h) \rangle^2 \, dh.$$

**Lemma 4.13.** *If $C(h) = \left(h, \frac{1}{2}h^2, \ldots, \frac{1}{m!}h^m\right)$ for $h \in [-1, 1]$, then $\delta(C, 1) > 0$.*

**Proof.** Suppose $\delta(C, 1) = 0$. Since the map $\eta \mapsto \int_{-1}^{1} \langle \eta, C(h) \rangle^2 \, dh$ is continuous on the unit sphere in $\mathbb{R}^m$, there must be some unit vector $\nu$ for which $\int_{-1}^{1} \langle \nu, C(h) \rangle^2 \, dh = 0$. Hence, $\langle \nu, C(h) \rangle = 0$ for all $h \in [-1, 1]$. However, $\nu \neq 0$, and so $\langle \nu, C(h) \rangle = \sum_{n=1}^{m} \frac{\nu_n}{n!} h^n$ is a nontrivial polynomial, a contradiction. □

**Lemma 4.14.** *Let $C(h)$ be as above, and let $Pert(h)$ be a continuous vector-valued function into $\mathbb{R}^m$ such that $\|Pert(h)\| \leq C_{Pert}|h|^{m+1}$ for $h \in [-1, 1]$. If $V(h) := C(h) + Pert(h)$, then $\delta(V, \epsilon) \geq \frac{1}{2}\delta(C, 1)\epsilon^{2m+1}$ as $\epsilon \to 0$.*

**Proof.** Note that $\langle \eta, V(h) \rangle^2 \geq \langle \eta, C(h) \rangle^2 + 2 \langle \eta, C(h) \rangle \langle \eta, Pert(h) \rangle$. Thus,

$$\begin{aligned} \delta(V, \epsilon) &\geq \min_{\|\eta\|=1} \int_{-\epsilon}^{\epsilon} \langle \eta, C(h) \rangle^2 \, dh + 2 \min_{\|\eta\|=1} \int_{-\epsilon}^{\epsilon} \langle \eta, C(h) \rangle \langle \eta, Pert(h) \rangle \, dh \\ &\geq \delta(C, \epsilon) - 2 \min_{\|\eta\|=1} \int_{-\epsilon}^{\epsilon} \Big| \langle \eta, C(h) \rangle \langle \eta, Pert(h) \rangle \Big| dh \end{aligned}$$

We bound this minimum as follows, where $e_m$ is the unit vector $(0, \ldots, 0, 1)$:

$$\begin{aligned} \min_{\|\eta\|=1} \int_{-\epsilon}^{\epsilon} \Big| \langle \eta, C(h) \rangle \langle \eta, Pert(h) \rangle \Big| dh &\leq \int_{-\epsilon}^{\epsilon} \Big| \langle e_m, C(h) \rangle \langle e_m, Pert(h) \rangle \Big| dh \\ &\leq \frac{C_{Pert}}{m!} \int_{-\epsilon}^{\epsilon} |h|^{2m+1} dh \\ &= \frac{C_{Pert}}{(m+1)!} \epsilon^{2m+2} \end{aligned}$$

Next, we determine a lower bound for $\delta(C, \epsilon)$. Let $E_\epsilon$ denote the diagonal matrix $E_\epsilon = diag(\epsilon, \epsilon^2, \ldots, \epsilon^m)$, and note that

$$\|E_\epsilon \eta\| \geq \epsilon^m \|\eta\| = \epsilon^m \qquad \text{for every unit vector } \eta.$$

Setting $h = \epsilon t$ and $\gamma = E_\epsilon \eta / \|E_\epsilon \eta\|$, we change variables in the integral defining $\delta(C, \epsilon)$:

$$\begin{aligned}\int_{-\epsilon}^{\epsilon} \langle \eta, C(h)\rangle^2 \, dh &= \epsilon \int_{-1}^{1} \langle E_\epsilon \eta, C(t)\rangle^2 \, dt \\ &= \epsilon \|E_\epsilon \eta\|^2 \int_{-1}^{1} \langle \gamma, C(t)\rangle^2 \, dt \\ &\geq \epsilon^{2m+1} \delta(C, 1)\end{aligned}$$

and taking the minimum over all unit vectors $\eta$, we obtain

$$\delta(C, \epsilon) \geq \delta(C, 1)\epsilon^{2m+1}.$$

Since $\delta(C, 1) > 0$, we have

$$\delta(V, \epsilon) \geq \delta(C, 1)\epsilon^{2m+1} - \frac{2C_{Pert}}{(m+1)!}\epsilon^{2m+2} \geq \frac{1}{2}\delta(C, 1)\epsilon^{2m+1}$$

for sufficiently small $\epsilon$. □

**Lemma 4.15.** *There is a constant $C_1 > 0$, independent of $\epsilon$, such that for all small $\epsilon > 0$*

$$\min_{\|\eta\|=1} \int_{p-\epsilon}^{p+\epsilon} \langle \eta, \pi(x) - P\rangle^2 \, dx \geq C_1 \epsilon^{2m+1}.$$

**Proof.** Recall that $\pi(x) - P = T(x-p) + Rem(x-p)$, where $T(x-p)$ is the order-$m$ Taylor polynomial:

$$T(x-p) = \sum_{n=1}^{m} \frac{1}{n!} \pi^{(n)}(p)(x-p)^n$$

and $Rem(x-p)$ is the Taylor remainder term satisfying

$$\|Rem(x-p)\| \leq C_{Taylor}|x-p|^{m+1}.$$

Let $Q$ be the change-of-basis matrix such that with respect to the canonical embedding basis, $Q^{-1}T(h) = \left(h, \frac{1}{2}h^2, \ldots, \frac{1}{m!}h^m\right)_P = C(h)$. Then, we can set $\gamma := Q^T\eta / \left\|Q^T\eta\right\|$ and $V(x-p) := C(x-p) + Q^{-1}Rem(x-p)$, and we have

$$\langle \eta, \pi(x) - P\rangle = \left\langle Q^T\eta, C(x-p) + Q^{-1}Rem(x-p)\right\rangle = \left\|Q^T\eta\right\| \langle \gamma, V(x-p)\rangle .$$

In order to apply Lemma 4.14 to $V$, we note that because $Q$ is a fixed nonsingular matrix, $\|Q^{-1}Rem(x-p)\| \leq \|Q^{-1}\| \, C_{Taylor}|x-p|^{m+1}$. We will also need to examine $\left\|Q^T\eta\right\|$. Observe that $Q^T$ acts on the unit circle to produce an ellipse

with axes of length $\sqrt{\sigma_i}$, where $\sigma_i > 0$ are the singular values of $Q$. Thus, for each unit vector $\eta$, we have $\left\|Q^T\eta\right\| \geq \sqrt{\sigma_{min}} > 0$ where $\sigma_{min} = \min \sigma_i$. Thus, setting $h = x - p$, we have for each unit vector $\eta$

$$\int_{p-\epsilon}^{p+\epsilon} \langle \eta, \pi(x) - P\rangle^2 \, dx = \left\|Q^T\eta\right\|^2 \int_{-\epsilon}^{\epsilon} \langle \gamma, V(h)\rangle^2 \, dh \geq \frac{1}{2}\sigma_{min}\delta(C,1)\epsilon^{2m+1}$$

by Lemma 4.14. The conclusion follows by taking the minimum over all unit vectors $\eta$. □

**Lemma 4.16.** *For each $m \times m$ matrix $E$, there is a unit vector $\eta_0 \in \mathbb{R}^m$ such that*

$$\|Ev\| \geq \|E\| \, |\langle \eta_0, v\rangle|$$

*for all vectors $v \in \mathbb{R}^m$.*

**Proof.** Let $\eta_0$ be a unit eigenvector for $E^TE$ corresponding to the largest eigenvalue $\alpha$ of $E^TE$. It follows from the theory of the singular value decomposition that $\alpha = \left\|E^TE\right\| = \|E\|^2$ [15, pp. 266]. Moreover, $E^TE$ is self-adjoint, hence diagonalizable with an orthonormal basis of eigenvectors. For any vector $v$ in $\mathbb{R}^m$, we can write $v = \langle \eta_0, v\rangle \, \eta_0 + w$ for some vector $w$ perpendicular to $\eta_0$. Since the various eigenspaces are orthogonal and invariant under $E^TE$, we have $\langle w, \eta_0\rangle = \left\langle E^TEw, \eta_0\right\rangle = \left\langle w, E^TE\eta_0\right\rangle = 0$. Therefore,

$$\begin{aligned}
\|Ev\|^2 &= \left\langle E^TEv, v\right\rangle \\
&= \langle \eta_0, v\rangle^2 \left\langle E^TE\eta_0, \eta_0\right\rangle + \left\langle E^TEw, w\right\rangle \\
&= \langle \eta_0, v\rangle^2 \langle \alpha\eta_0, \eta_0\rangle + \|Ew\|^2 \\
&\geq \langle \eta_0, v\rangle^2 \, \alpha
\end{aligned}$$

The conclusion follows by taking square roots. □

**Proof of Theorem 4.11.** The hypotheses guarantee that the Eckmann-Ruelle linearization $M = M(P)$ exists and is the unique best linearization as $\epsilon_k \to 0$. We prove that $\|M_k - M\| \to 0$ as $k \to \infty$. For any $k$, we have

$$\begin{aligned}
&\int_{I(P,\epsilon_k)} \|(M_k - M)\,(\pi(x) - P)\|^2 \, d\mu(x) \\
&\qquad \leq \int_{I(P,\epsilon_k)} \left(\Phi_{M_k}(x, \pi^{-1}(P))^{\frac{1}{2}} + \Phi_M(x, \pi^{-1}(P))^{\frac{1}{2}}\right)^2 d\mu(x) \\
&\qquad \leq \left(LS(M_k, P, \epsilon_k) + LS(M, P, \epsilon_k)\right)^2
\end{aligned}$$

where we have used the Hölder Inequality for the last step. By hypothesis and (4.10), it follows that

$$\int_{I(P,\epsilon_k)} \|(M_k - M)(\pi(x) - P)\|^2 \, d\mu(x) \leq 4C^2\epsilon_k^{2m+3}.$$

On the other hand, by Lemma 4.16, there is a unit vector $\eta_0$ such that

$$\int_{I(P,\epsilon_k)} \|(M_k - M)(\pi(x) - P)\|^2 \, d\mu(x) \\ \geq \|M_k - M\|^2 \int_{I(P,\epsilon_k)} \langle \eta_0, \pi(x) - P\rangle^2 \, d\mu(x).$$

We need a lower bound in terms of $\epsilon$ for the integral on the right. Note that the interval $I(P,\epsilon_k)$ contains the interval $\left[p - \frac{\epsilon_k}{C_\pi}, p + \frac{\epsilon_k}{C_\pi}\right]$. Using the continuity and positivity of the density function $\mathcal{F}$ and Lemma 4.15, we have for sufficiently small $\epsilon_k$:

$$\begin{aligned}\int_{I(P,\epsilon_k)} \langle \eta_0, \pi(x) - P\rangle^2 \, d\mu(x) &\geq \frac{1}{2}\mathcal{F}(p) \int_{p-\frac{\epsilon_k}{C_\pi}}^{p+\frac{\epsilon_k}{C_\pi}} \langle \eta_0, \pi(x) - P\rangle^2 \, dx \\ &\geq \frac{1}{2}\mathcal{F}(p) C_1 \left(\frac{\epsilon_k}{C_\pi}\right)^{2m+1}\end{aligned}$$

for some constant $C_1 > 0$, independent of $\epsilon_k$. It follows that

$$\|M_k - M\|^2 \, \frac{\mathcal{F}(p)C_1}{2C_\pi^{2m+1}} \, \epsilon_k^{2m+1} \leq 4C^2\epsilon_k^{2m+3}$$

and therefore $\|M_k - M\| = O(\epsilon_k)$ as $\epsilon_k \to 0$, completing the proof. □

# Chapter 5

# Numerical Computations

In this chapter, we describe numerical experiments illustrating the theory developed in the previous chapters. As we shall see, we can compute numerically the Eckmann-Ruelle linearization in (noise-free) examples. The algorithm we use is similar to that outlined in [2], though the specific program was written to handle any function as a measurement function not just time-delay measurement functions. We briefly outline the algorithm here.

Recall that the Eckmann-Ruelle procedure has three basic steps. In the first step, one reconstructs the attractor with a measurement function. In the second step, one determines a local linearization matrix at each point of the reconstructed trajectory. In the final step, one extracts the Lyapunov exponents from the linearization matrices obtained in the previous step.

For our numerical experiments, we assume that we are given the underlying dynamical system $f : \mathbb{R}^n \to \mathbb{R}^n$ and a measurement function $\pi : \mathbb{R}^n \to \mathbb{R}^m$ for reconstructing the attractor. We iterate the function $f$ a number of times to remove transients. Then, continuing to iterate $f$, we produce a trajectory $\{p_i\} \in \mathbb{R}^n$. Instead of storing the points $p_i$, we apply the measurement function to each $p_i$ and store the vectors $P_i = \pi(p_i) \in \mathbb{R}^m$. We then sort our list of $P_i$ by their first coordinates, keeping track of where each $P_i$ ends up in the sorted list.

The procedure for computing the local linearization matrix at a base point $P$ is straightforward and based on least-squares methods. The goal is to determine the $m \times m$ matrix $M$ which best satisfies $M(P_i - P) \approx P_{i+1} - F(P)$ for all $P_i$ close to $P$. Given a neighborhood radius $\epsilon$, we quickly search our sorted list for those $P_i$ whose first coordinate lies within a distance $\epsilon$ of the first coordinate of $P$. Any other point $P_i$ certainly lies outside the ball in $\mathbb{R}^m$ of radius $\epsilon$ about $P$. This first search often reduces the number of points we need to check carefully to less than 10% of the trajectory. For each $P_i$ in our shortened list, we construct the vectors $P_i - P$ and $P_{i+1} - F(P)$ and check their lengths. If both of these vectors have length less than or equal to $\epsilon$, then we keep them. Since the linearization we seek satisfies $M\,(P_i - P) \approx P_{i+1} - F(P)$, we build matrices $A$ and $B$ for which the $k$-th column of $A$ is a vector $P_i - P$ and the $k$-th column of $B$ is the corresponding

vector $P_{i+1} - F(P)$. Then, our local linearization matrix $M$ will be the $m \times m$ solution of the matrix equation $MA = B$. We solve this equation using a singular value decomposition routine taken from [16].

Once the local linearization matrices have been computed, we must compute the exponents. Recall from Chapter 3 that the Eckmann-Ruelle-Lyapunov (ERL) exponents of the reconstructed trajectory $P_0, P_1, \ldots$ in $\mathbb{R}^m$ are the values obtained by the limit

$$h_{ER}(P_0, \nu) := \lim_{n\to\infty} \frac{1}{n} \ln \|M_{n-1} M_{n-2} \cdots M_1 M_0 \nu\|$$

for unit vectors $\nu \in \mathbb{R}^m$, where $M_i = M(P_i)$ is the $m \times m$ local linearization matrix (i.e., the Eckmann-Ruelle linearization) at the point $P_i$ in the trajectory. To evaluate this limit, we employ the treppen-iteration algorithm suggested in [1, 2]. Given our sequence $M_i$ of linearization matrices, we use the QR matrix decomposition to find orthogonal matrices $Q_i$ and upper-triangular matrices $R_i$ (with non-negative diagonal elements) such that

$$M_i Q_{i-1} = Q_i R_i \qquad \text{for } i = 0, 1, 2, \ldots$$

where we take $Q_{-1}$ to be the $m \times m$ identity matrix. Then, we can write

$$M_{n-1} M_{n-2} \cdots M_1 M_0 = Q_{n-1} R_{n-1} R_{n-2} \cdots R_0.$$

The orthogonal matrix $Q_{n-1}$ won't affect the matrix norm. Thus,

$$h_{ER}(P_0, \nu) = \lim_{n\to\infty} \frac{1}{n} \ln \|R_{n-1} R_{n-2} \cdots R_1 R_0 \nu\| .$$

The product $R_{n-1} \cdots R_0$ is upper-triangular, and its diagonal elements are the eigenvalues of the matrix, which we expect will grow like the Lyapunov exponents. As mentioned in Chapter 3, Theorem 3.2 justifies reading the exponents directly from the diagonal:

$$\lambda_k = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln\left((R_i)_{kk}\right) \qquad \text{for } k = 1, \ldots, m.$$

Thus, we are able to compute the ERL exponents from our linearizations by using the diagonal elements from the QR decomposition.

We shall now discuss several examples.

In Chapter 2, we considered the doubling map on an interval reconstructed on the unit circle in $\mathbb{R}^2$ where the underlying dynamics was $f(p) = 2p \ (mod\ 2\pi)$ on $[0, 2\pi]$, and the measurement function was $\pi(p) = (\cos(p), \sin(p))$. We computed the Eckmann-Ruelle linearization in this case to be

$$M(x, y) = \begin{pmatrix} 4x^3 & -4y^3 \\ 2y\,(2x^2 + 1) & 2x\,(2y^2 + 1) \end{pmatrix} \tag{5.1}$$

as a function of the points $(x, y)$ on the unit circle. In numerical experiments, our routines to compute the best local linearization find good approximations to this matrix. For example, we can compute the linearization matrix at the fixed point (1,0) of the reconstructed dynamics $F$. Using a trajectory of 100,000 data points, the linearization matrix for a neighborhood of radius $\epsilon = 0.05$ is computed to be

$$\begin{pmatrix} 3.999601 & 0.000000 \\ 0.001048 & 1.999675 \end{pmatrix} \approx \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} = M(1, 0).$$

Theorems 4.2 and 4.11 suggest that as the neighborhood radius decreases, the computed linearizations should converge in matrix norm to $M(1, 0)$. Figure 5.1 shows a graph of this convergence as the radius shrinks to zero. To generate this graph, we first compute and store a reconstructed trajectory of 100,000 data points on the unit circle. Then, for each radius $\epsilon$, we find those points in the trajectory within the specified radius of our base point and use them to compute the linearization matrix.

Likewise, the local linearizations converge at other points on the unit circle. For a neighborhood radius of $\epsilon = 0.05$, the linearization matrix computed at $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \approx (.707, .707)$ is

$$\begin{pmatrix} 1.414398 & -1.413499 \\ 2.828112 & 2.828112 \end{pmatrix} \approx \begin{pmatrix} \sqrt{2} & -\sqrt{2} \\ 2\sqrt{2} & 2\sqrt{2} \end{pmatrix} = M\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right).$$

Figure 5.2 shows the convergence with radius of the linearization matrices at this point.

For a given neighborhood radius, we can determine at each point on the circle the error between the computed linearization and the Eckmann-Ruelle linearization. Figure 5.3 shows this graph for several different radii.

With close agreement at each point in the interval between the numerically-determined linearization and the Eckmann-Ruelle linearization, we expect that the computed Lyapunov exponents will match those given by the Lyapunov exponent formula of Theorem 3.3. To test this, we computed the Lyapunov exponents for this example using 100,000 data points on the circle. We found values of $0.693173 \approx \ln(2)$ and $1.386244 \approx 2\ln(2)$, as predicted. The first value is the true Lyapunov exponent of the doubling map, while the second value is a spurious exponent. Notice that the largest computed value is not a Lyapunov exponent of the underlying dynamical system.

The Lyapunov exponent formula of Theorem 3.3 holds for any generic map and specifically for the one-to-one maps with linearly independent derivative vectors. Since the formula itself is independent of the measurement function $\pi$, we expect to compute the same exponents when different measurement functions are used. Changing the measurement function will likely change the speed of convergence somewhat, but we still expect the results to be close. With this in mind, we

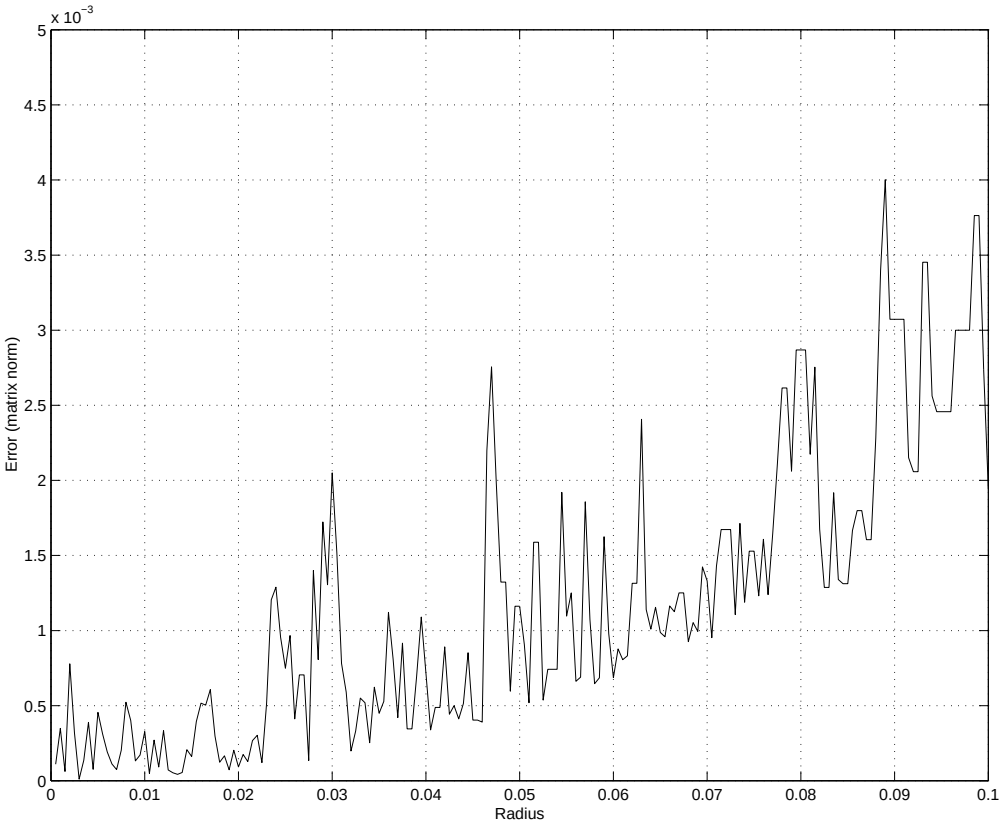



Figure 5.1: Graph of the difference (in matrix norm) between the computed linearization and the Eckmann-Ruelle linearization at the fixed point (1,0) for the doubling map $f(p) = 2p(mod\ 2\pi)$ reconstructed on the unit circle. The calculation used 100,000 data points. Note the downward trend of the graph, indicating the convergence of the linearizations to the Eckmann-Ruelle matrix. The spikes in the graph appear to be numerical artifacts. For radii below about 0.01, very few data points lie sufficiently close to the base point to use in the calculation.

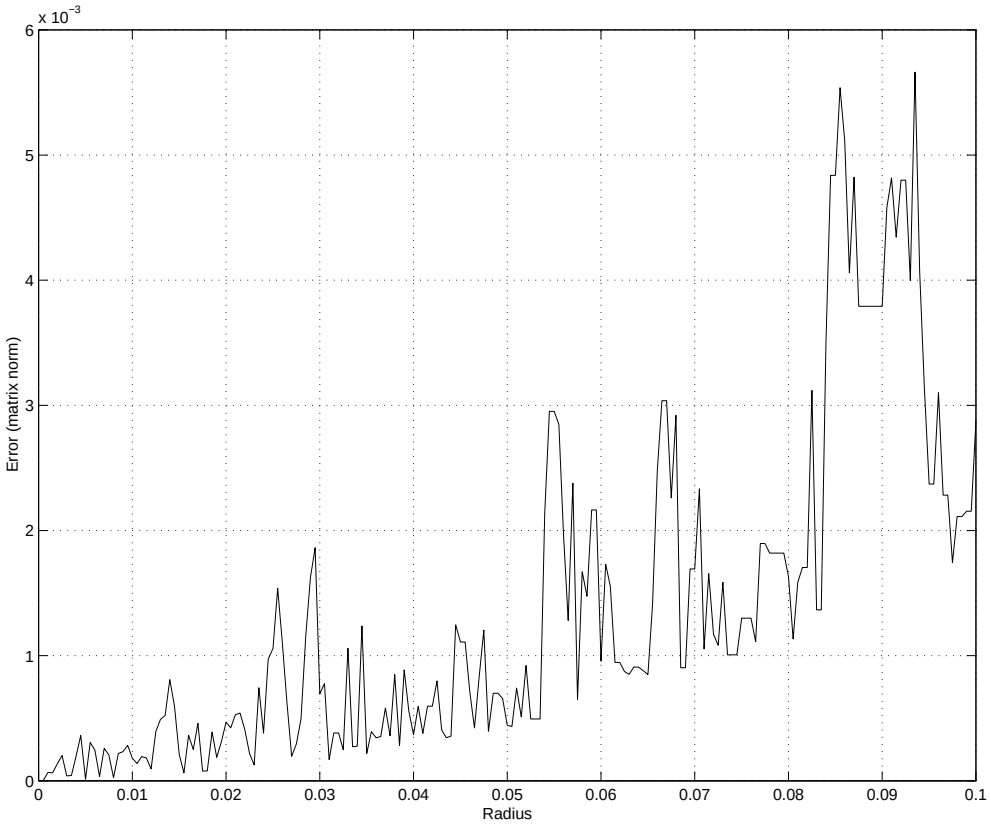



Figure 5.2: Graph of the difference (in matrix norm) between the computed linearization and the Eckmann-Ruelle linearization at the point $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ for the doubling map reconstructed on the unit circle. The calculation used 100,000 data points. Note the downward trend of the graph, indicating the convergence of the linearizations to the Eckmann-Ruelle matrix. The spikes in the graph appear to be numerical artifacts. For radii below about 0.01, very few data points lie sufficiently close to the base point to use in the calculation.

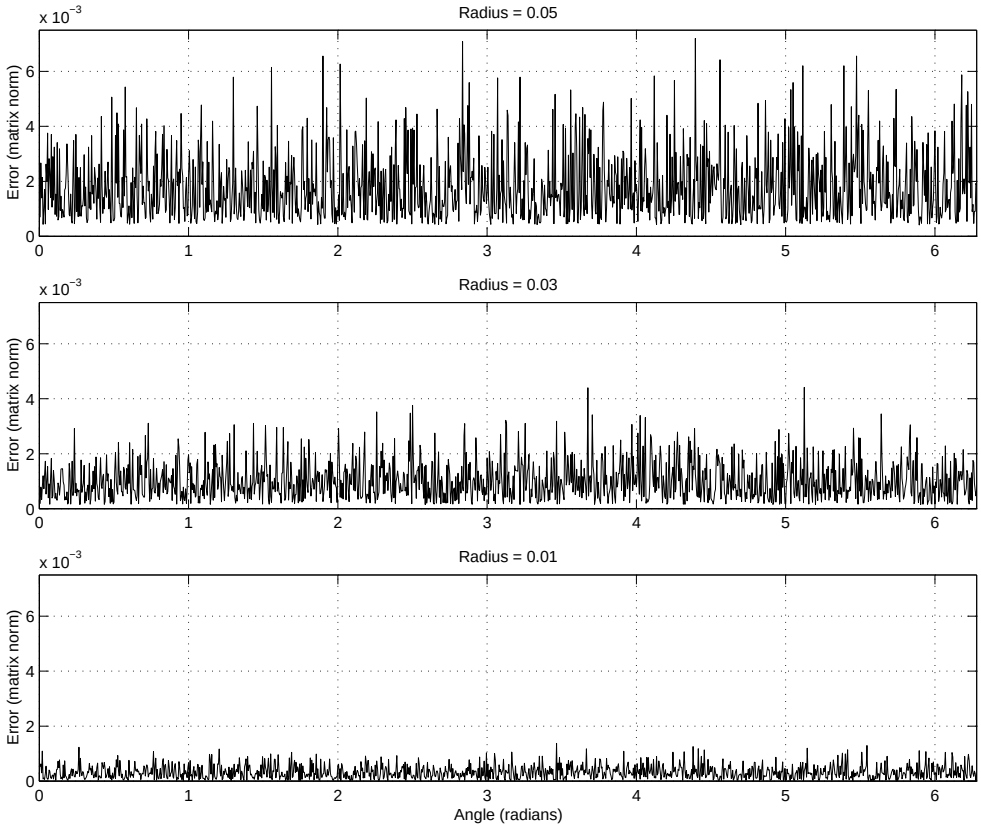



Figure 5.3: These graphs show, for different radii, the error in matrix norm between the computed linearization and the Eckmann-Ruelle linearization for the doubling map at each point on the unit circle (expressed as an angle). As the radius decreases, the error decreases uniformly. The graphs were created using a single trajectory of 200,000 data points on the unit circle.

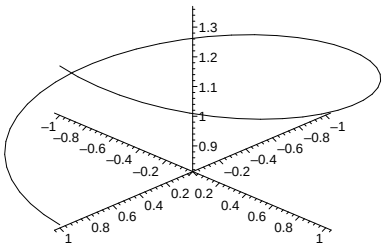



Figure 5.4: The curve $\pi(p) = (\cos(2\pi p), \sin(2\pi p), \frac{1}{2}e^{1-\frac{1}{2}p})$ for $p \in [0, 1]$.

performed a Lyapunov exponent calculation for the doubling map reconstructed on an ellipse using the measurement function $\pi(p) = (2\cos(p), \sin(p) - \frac{1}{2}\cos(p))$ and found exponents of $0.693151 \approx \ln(2)$ and $1.386267 \approx 2\ln(2)$.

Next, we examine the case of a reconstruction from $\mathbb{R}^1$ into $\mathbb{R}^3$. We use the logistic map $f(p) = 4p(1-p)$ on $[0, 1]$ reconstructed onto a curve by the measurement function $\pi(p) = (\cos(2\pi p), \sin(2\pi p), \frac{1}{2}e^{1-\frac{1}{2}p})$. Figure 5.4 shows this curve in $\mathbb{R}^3$.

In this case, the formula for the Eckmann-Ruelle linearization is more complicated than (5.1), but for any individual base point, it is easily computed from the definition given in Chapter 2. For example, at the point $\pi(0.125) \approx (0.707, 0.707, 1.277)$ on the curve in $\mathbb{R}^3$, the Eckmann-Ruelle linearization is

$$M(\pi(0.125)) = \begin{pmatrix} -0.395832 & 2.313014 & -18.147168 \\ -6.279704 & -3.080283 & 65.170183 \\ -0.000276 & -0.098277 & 1.550445 \end{pmatrix}$$

and, computing with neighborhood radius $\epsilon = 0.05$, the local linearization matrix is

$$\begin{pmatrix} -0.391945 & 2.313262 & -18.187396 \\ -6.273483 & -3.079227 & 65.103345 \\ -0.000214 & -0.098268 & 1.549787 \end{pmatrix}.$$

Figure 5.5 shows the error in the linearizations over the whole curve. To compute the Lyapunov exponents for this example, we used 100,000 data points on the curve. We found values $0.6884401 \approx \ln(2)$, $1.392024 \approx 2\ln(2)$, and $2.056767 \approx 3\ln(2)$. This is consistent with Theorem 3.3 since the Lyapunov exponent for the logistic map is $\ln(2)$. Note that the largest computed exponent is spurious.

We also computed the Lyapunov exponents of several other reconstructions of the logistic map in $\mathbb{R}^3$. The results are listed in Table 5.1.

In addition to the somewhat arbitrary measurement functions we have used thus far, we can also construct time-delay embeddings in $\mathbb{R}^3$ of the logistic map. At each iteration, we record some quantity based on the current state of the system. In an actual experiment, this choice would be realized by the appara-

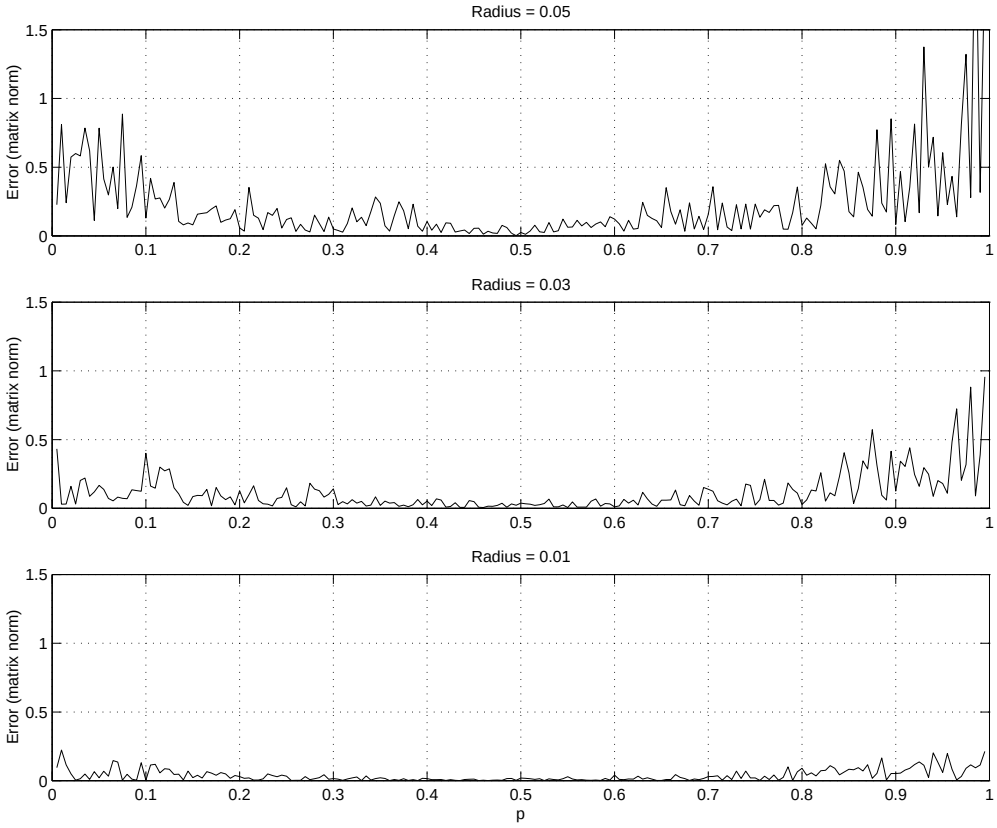



Figure 5.5: These graphs show, for different radii, the error in matrix norm between the computed linearization and the Eckmann-Ruelle linearization of the logistic map at each point on the curve in $\mathbb{R}^3$. As the radius decreases, the error decreases uniformly. The graphs were created using a single trajectory of 200,000 data points on the curve in $\mathbb{R}^3$.

Table 5.1: The Eckmann-Ruelle-Lyapunov exponents of several reconstructions of the logistic map $f(p) = 4p(1-p)$ into $\mathbb{R}^3$. Each computation involved 100,000 data points. The computed exponents are roughly $\ln(2)$, $2\ln(2)$, and $3\ln(2)$.

| Measurement Function $\pi(p)$ | Computed ERL Exponents | | |
|---|---|---|---|
| $(p, p^2, p^3)$ | 0.664767 | 1.411537 | 2.082149 |
| $(p, \sqrt{p} - 2p^3, \frac{1}{2}p\cos(2.5\pi p))$ | 0.678497 | 1.391379 | 2.075309 |
| $(\cos(2\pi p), p + \sin(2\pi p), p + p\sin(5\pi p^2))$ | 0.687375 | 1.385631 | 2.066531 |

Table 5.2: The Eckmann-Ruelle-Lyapunov exponents of several time-delay reconstructions of the logistic map $f(p) = 4p(1-p)$ into $\mathbb{R}^3$. For each iteration, we recorded a scalar determined by the current state $p$ of the system. Each computation used 100,000 data points. The computed exponents are roughly $\ln(2)$, $2\ln(2)$, and $3\ln(2)$.

| Recorded Quantity | Computed ERL Exponents | | |
|---|---|---|---|
| $p$ | 0.691530 | 1.386031 | 2.083390 |
| $\cos(2p) - p$ | 0.692866 | 1.386281 | 2.080733 |
| $\ln(1+p)$ | 0.691398 | 1.386725 | 2.079474 |

tus used to measure and record the data. The results of computing Lyapunov exponents for several time-delay examples are given in Table 5.2.

The Eckmann-Ruelle linearization described in (5.1) may be misleadingly simple. It is possible to compute an explicit formula for the Eckmann-Ruelle linearization for the case of the logistic map $f(p) = 4p(1-p)$ on $[0,1]$ reconstructed onto the unit circle by the measurement function $\pi(p) = (\cos(2\pi p), \sin(2\pi p))$. One then finds that some of the component functions of that matrix can have derivatives as large as 200 in absolute value. See Figure 5.6.

We move on to discuss the case of two-dimensional underlying dynamics. Here, we base our experiments on the Hénon map $f(x,y) = (1.4 - x^2 + 0.3y, x)$. We begin by considering the measurement function $\pi(x,y) = (x, y, x^2, y^2, xy)$. For this measurement function, we can compute the Eckmann-Ruelle linearization explicitly:

$$M(\pi(x,y)) = \begin{pmatrix} 0 & 0.3 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1.2xy - 8x^3 & 0.84 + 0.6x^2 & -2.8 + 6x^2 - 0.6y & .09 & -1.2x \\ 0 & 0 & 1 & 0 & 0 \\ 1.4 + 3x^2 & 0 & -3x & 0 & 0.3 \end{pmatrix}$$

where for convenience, we refer to $M$ using coordinates in the underlying space $\mathbb{R}^2$ instead of coordinates in the reconstruction space $\mathbb{R}^5$. We observe the same phenomena when the underlying dynamics are two-dimensional as we did in the previous cases. For example, at the point $\pi(1.555478, 0.398567) \in \mathbb{R}^5$ on the reconstructed Hénon attractor, the linearization matrix for a neighborhood of radius $\epsilon = 0.01$ is computed to be

$$\begin{pmatrix} 0.000000 & 0.300000 & -1.000000 & 0.000000 & 0.000000 \\ 1.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ -29.365725 & 2.290703 & 11.478416 & 0.090305 & -1.866085 \\ 0.000000 & 0.000000 & 1.000000 & 0.000000 & 0.000000 \\ 8.657187 & -0.001385 & -4.666094 & 0.000326 & 0.300723 \end{pmatrix}$$

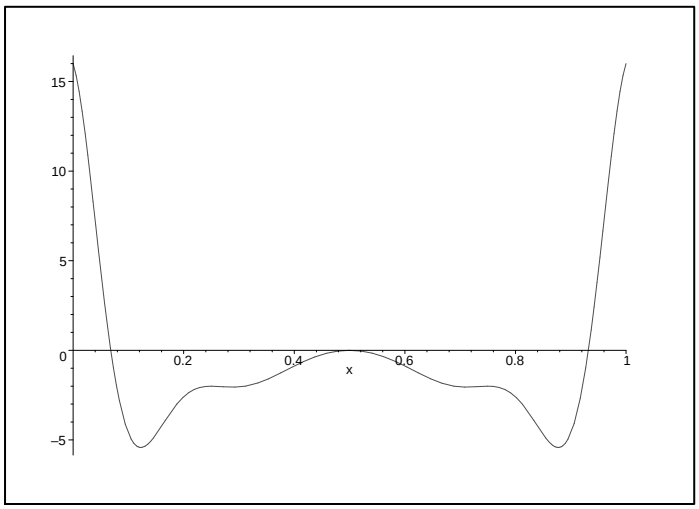



Figure 5.6: A graph of the first coordinate of the Eckmann-Ruelle matrix for the logistic map $f(p) = 4p(1-p)$ on [0,1] reconstructed on the unit circle in $\mathbb{R}^2$ by the measurement function $\pi(p) = (\cos(2\pi p), \sin(2\pi p))$. Note the large derivatives near the ends of the interval.

At another point, $\pi(-1.741541, 1.753985)$, on the reconstructed Hénon attractor in $\mathbb{R}^5$, the local linearization for $\epsilon = 0.01$ is computed to be

$$\begin{pmatrix} 0.000000 & 0.300000 & -1.000000 & 0.000000 & 0.000000 \\ 1.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ 38.590278 & 2.658478 & 14.345224 & 0.090317 & 2.089740 \\ 0.000000 & 0.000000 & 1.000000 & 0.000000 & 0.000000 \\ 10.498785 & -0.000234 & 5.224592 & 0.000067 & 0.300000 \end{pmatrix}.$$

Both of these examples are in good agreement with the general formula above. As before, we can graph the convergence of the local linearizations to the Eckmann-Ruelle linearization as the neighborhood radius shrinks to zero. See Figure 5.7. Using 300,000 data points, we computed the Lyapunov exponents for this example. Recall that the true Lyapunov exponents of the Hénon map are approximately $\lambda = 0.42$ and $\mu = -1.62$. We found values of $0.886200 \approx 2\lambda$, $0.422459 \approx \lambda$, $-1.195450 \approx \lambda + \mu$, $-1.636780 \approx \mu$, and $-3.183227 \approx 2\mu$. These values are consistent with the Lyapunov exponent formula given in Theorem 3.6.

We also computed the Lyapunov exponents from time-delay reconstructions of the Hénon system. Table 5.3 shows the results. As one can see in Table 5.3, we do not find all of the exponents predicted by the Lyapunov exponent formula in Theorem 3.6. Three of the exponents match nicely with the predicted formula, namely $\lambda \approx 0.42$, $2\lambda \approx 0.84$, and $\mu \approx -1.62$, but the other two vary somewhat from their expected values. With more data and smaller neighborhood radii in

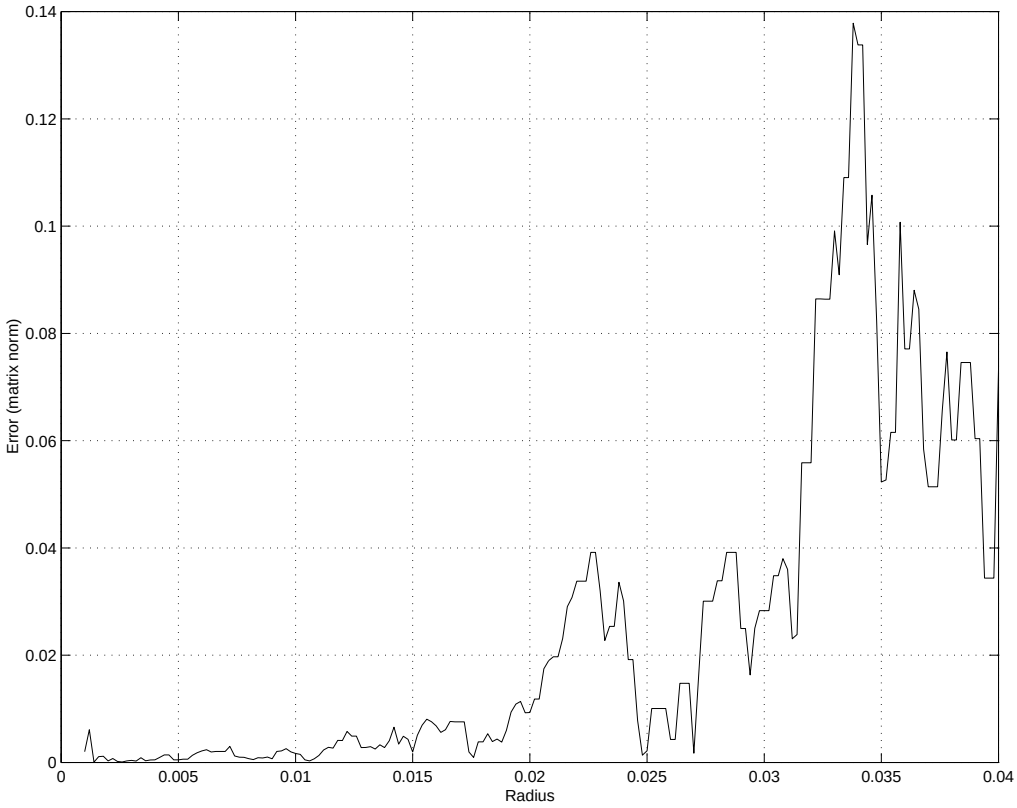



Figure 5.7: Graph of the difference (in matrix norm) between the computed linearization and the Eckmann-Ruelle linearization at the point (-1.741541, 1.753985) for the Hénon map reconstructed in $\mathbb{R}^5$. The calculation used 1,000,000 data points. The graph does not show radii larger than about 0.04 because this base point lies about 0.045 units away from a significant bend in the Hénon attractor.

Table 5.3: The Eckmann-Ruelle-Lyapunov exponents of several time-delay reconstructions of the Hénon map into $\mathbb{R}^5$. For each iteration, we recorded a scalar determined by the current state $(x, y)$ of the system. Each computation involved 300,000 data points. Three of the computed exponents match the Lyapunov exponent formula: $0.84 \approx 2\lambda$, $0.42 \approx \lambda$, and $-1.62 \approx \mu$. The other exponents will probably converge for longer time series.

| Recorded Quantity | Computed ERL Exponents | | | | |
|---|---|---|---|---|---|
| $x$ | 0.838426 | 0.418513 | -1.070009 | -1.619577 | -2.288856 |
| $y - xy$ | 0.839063 | 0.418518 | -1.064124 | -1.619594 | -2.426907 |
| $\frac{1}{2}(x^2 + y^2)$ | 0.837956 | 0.418331 | -0.846145 | -1.621547 | -2.404775 |
| $\arctan\left(\frac{y}{x}\right)$ | 0.840121 | 0.418186 | -1.062045 | -1.618537 | -2.226276 |
| $\cos(x) + \sin(y)$ | 0.839557 | 0.418119 | -1.019653 | -1.621305 | -2.378717 |
| $\cos(y-1) - \frac{3}{4}x^2$ | 0.839295 | 0.418780 | -1.095620 | -1.621523 | -2.624901 |

the computations, those last two exponents would likely converge to their proper values. Note that the true Lyapunov exponents appear in each of these examples, suggesting that they converge to their correct values rather quickly. This is not surprising if one looks back at the proofs of Theorem 2.6 in Chapter 2 and Theorem 3.6 in Chapter 3. Any matrix that does not map the first-order terms of the Taylor series correctly will have $O(\|\Delta P\|)$ error, instead of the $O(\|\Delta P\|^3)$ error of the Eckmann-Ruelle linearization. Fortunately, algorithms to compute local linearizations will instead find matrices with error $O(\|\Delta P\|^2)$ or better, and these matrices will map the first-order Taylor terms correctly. This ensures that the true Lyapunov exponents will be among the first of the computed exponents to converge.

# Appendix A

# Polynomial Results

In this appendix, we derive the facts about polynomials that we need in the text.

**Proposition A.1.** *Let $p(x)$ be a polynomial of degree $d \geq 0$, and suppose there are infinitely many values $x_i \to 0$ for which $|p(x_i)| \leq C|x_i|^q$ for some $q > d$. Then $p(x) = 0$ for all $x$.*

**Proof.** Write $p(x) = a_d x^d + \cdots + a_1 x + a_0$. By the continuity of $p(x)$,

$$0 \leq |a_0| = |p(0)| = \lim_{x_i \to 0} |p(x_i)| \leq \lim_{x_i \to 0} C|x_i|^q = 0$$

since $q > d \geq 0$. Thus, $a_0 = 0$.

For induction, suppose that $a_0 = a_1 = \cdots = a_{k-1} = 0$ where $k \leq d < q$. Write $p(x) = x^k(a_d x^{d-k} + \cdots + a_{k+1}x + a_k) := x^k q_k(x)$. Then, by the continuity of $q_k(x)$,

$$0 \leq |a_k| = |q_k(0)| = \lim_{x_i \to 0} |q_k(x_i)| = \lim_{x_i \to 0} \frac{|p(x_i)|}{|x_i|^k} \leq \lim_{x_i \to 0} \frac{C|x_i|^q}{|x_i|^k} = \lim_{x_i \to 0} C|x_i|^{q-k} = 0$$

since $q - k \geq q - d > 0$. Thus, $a_k = 0$, and the lemma follows from induction on $k$. □

We would like to prove a similar result for polynomials with more variables. Unfortunately, the result is not true, even for two variables, without extra hypotheses. For example, if $p(x, y) = x - y$, then $p(x_i, x_i) = 0$ for any sequence $x_i \to 0$. Thus, $p(x, y)$ need not be zero, even though there is a sequence going to $(0, 0)$ satisfying $|p(x_i, y_i)| \leq C\|(x_i, y_i)\|^q$ for every $q > 0$. The added hypotheses in the next lemma guarantee that the polynomial must be small in many directions, not just one unlucky choice.

Recall the definition of an approach direction from our discussion in Chapter 2 of the case where the underlying dynamics is two-dimensional.

**Definition A.2.** *Let $l$ be a unit vector in $\mathbb{R}^m$. A subset $S$ of $\mathbb{R}^m$ has the* **approach direction** *$l$ at the base point $P \in \mathbb{R}^m$ provided there is a sequence $\{Q_k\}_{k=1}^{\infty}$ from $S$ such that:*

1. *$Q_k \to P$ as $k \to \infty$, and*
2. *$\frac{\Delta Q_k}{\|\Delta Q_k\|} \to l$ as $k \to \infty$, where $\Delta Q_k := Q_k - P$.*

*We call a collection of approach directions at $P$* **distinct** *provided that no two are the same and no two are reflections through the origin. Multiple approach directions are not required to be linearly independent.*

**Proposition A.3.** *Let $p(x, y)$ be a polynomial of (total) degree $d$ (that is, each monomial in $p$ has total degree at most $d$). Assume that there is a sequence of points $(x_k, y_k) \to (0, 0)$ in $\mathbb{R}^2$ such that $|p(x_k, y_k)| \leq C\|(x_k, y_k)\|^q$ for some $q > d$. If the set $\{(x_k, y_k) : k \geq 0\}$ has $d + 1$ distinct approach directions at (0,0), then $p(x, y) = 0$ for all $(x, y) \in \mathbb{R}^2$.*

**Proof.** Write $p(x, y) = \sum_{j=0}^{d} p_j(x, y)$, where $p_j(x, y) = \sum_{e=0}^{j} a_{je} x^e y^{j-e}$ is a polynomial of degree $j$ with only monomials of total degree $j$. We will prove inductively that each polynomial $p_l(x, y)$ is identically zero for $l = 0, 1, \ldots, d$.

We begin with the $l = 0$ case. Note that

$$0 \leq |a_{00}| = |p(0, 0)| = \lim_{k\to\infty} |p(x_k, y_k)| \leq \lim_{k\to\infty} C\|(x_k, y_k)\|^q = 0$$

since $q > d \geq 0$. Thus, $p_0(x, y) = a_{00} = 0$, and so $p$ has no constant term.

Without loss of generality, we may now assume that $d \geq 1$.

Next, we show that $p_1(x, y) = a_{11}x + a_{10}y = 0$ by proving that $a_{11} = a_{10} = 0$. This is the $l = 1$ case and it will demonstrate the basic idea for the induction that follows. For an approach direction $(\alpha, \beta)$, let $\{(\hat{x}_k, \hat{y}_k)\}$ be a subsequence such that $(\hat{x}_k, \hat{y}_k)/\|(\hat{x}_k, \hat{y}_k)\| \to (\alpha, \beta)$. For convenience, set $n_k := \|(\hat{x}_k, \hat{y}_k)\|$. Then,

$$\begin{aligned}
0 \leq |a_{11}\alpha + a_{10}\beta| &= \lim_{k\to\infty} \left| a_{11}\frac{\hat{x}_k}{n_k} + a_{10}\frac{\hat{y}_k}{n_k} \right| = \lim_{k\to\infty} \frac{|p_1(\hat{x}_k, \hat{y}_k)|}{n_k} \\
&\leq \limsup_{k\to\infty} \left( \frac{|p(\hat{x}_k, \hat{y}_k)|}{n_k} + \sum_{j=2}^{d} \sum_{e=0}^{j} |a_{je}| \frac{|\hat{x}_k|^e |\hat{y}_k|^{j-e}}{n_k} \right) \\
&\leq \limsup_{k\to\infty} \left( \frac{Cn_k^q}{n_k} + \sum_{j=2}^{d} |a_{jj}| |\hat{x}_k|^{j-1} \frac{|\hat{x}_k|}{n_k} + \sum_{j=2}^{d} \sum_{e=0}^{j-1} |a_{je}| |\hat{x}_k|^e |\hat{y}_k|^{j-e-1} \frac{|\hat{y}_k|}{n_k} \right) \\
&= 0
\end{aligned}$$

because $q > d \geq 1$, $|\hat{x}_k|/n_k \to \alpha$, $|\hat{y}_k|/n_k \to \beta$, and $|\hat{x}_k|, |\hat{y}_k| \to 0$. Therefore, $a_{11}\alpha + a_{10}\beta = 0$ for each approach vector $(\alpha, \beta)$. When $\alpha \neq 0$, we can divide by $\alpha$ to get:

$$a_{11} + a_{10}\left(\frac{\beta}{\alpha}\right) = 0$$

We must now consider several cases.

CASE 1: *Assume that either $d \geq 2$, or $d = 1$ but there are 2 distinct approach vectors with $\alpha \neq 0$.* In this case, there are at least two approach vectors $(\alpha, \beta)$ with $\alpha \neq 0$. If we write $f(t) = a_{10}t + a_{11}$, then for each such vector, we have $f\left(\frac{\beta}{\alpha}\right) = 0$. As there are at least two such values and $f(t)$ is linear, we must have $f(t)$ identically zero. Hence, $a_{11} = a_{10} = 0$ and therefore $p_1(x, y) = 0$ for all $(x, y)$.

CASE 2: *Assume that $d = 1$ and that one of the 2 approach directions is $(0, \pm 1)$.* In this case, we see directly that $|a_{10}| = |a_{11}0 + a_{10}(\pm 1)| = 0$. Then, for the approach vector $(\alpha, \beta)$ with $\alpha \neq 0$, we have $0 = |a_{11}\alpha + 0 \cdot \beta| = |a_{11}\alpha|$. We conclude that $a_{11} = 0$ as well. Therefore $p_1(x, y) = 0$ for all $(x, y)$.

In all cases, $p_1(x, y) = 0$ as desired. The main induction step that follows is similar.

Suppose we have shown that $p_0(x, y) = p_1(x, y) = \cdots = p_{l-1}(x, y) = 0$ are all identically zero for some $l \leq d$. We want to show that $p_l(x, y) = 0$ is also identically zero. For any approach direction $(\alpha, \beta)$ (using the same notation as above):

$$\begin{aligned}
0 \leq \left|\sum_{e=0}^{l} a_{le}\alpha^e\beta^{l-e}\right| &= \lim_{k\to\infty}\left|\sum_{e=0}^{l} a_{le}\left(\frac{\hat{x}_k}{n_k}\right)^e\left(\frac{\hat{y}_k}{n_k}\right)^{l-e}\right| = \lim_{k\to\infty}\frac{|p_l(\hat{x}_k, \hat{y}_k)|}{n_k^l} \\
&\leq \limsup_{k\to\infty}\left(\frac{|p(\hat{x}_k, \hat{y}_k)|}{n_k^l} + \sum_{j=l+1}^{d}\frac{|p_j(\hat{x}_k, \hat{y}_k)|}{n_k^l}\right) \\
&\leq \limsup_{k\to\infty}\left(Cn_k^{q-l} + \sum_{j=l+1}^{d}\sum_{e=0}^{j} T_{je}^k\right) \\
&= 0
\end{aligned}$$

where the terms $T_{je}^k$ look like

$$T_{je}^k = |a_{je}||\hat{x}_k|^{e-p_1}|\hat{y}_k|^{j-e-p_2}\left(\frac{|\hat{x}_k|}{n_k}\right)^{p_1}\left(\frac{|\hat{y}_k|}{n_k}\right)^{p_2}$$

for some appropriate choice of $p_1 + p_2 = l < j$. The lim sup above is zero because

$$q - l \geq q - d > 0, \quad \frac{1}{n_k}(\hat{x}_k, \hat{y}_k) \to (\alpha, \beta), \quad \text{and } |\hat{x}_k|, |\hat{y}_k| \to 0$$

together imply that for each $j$ and $e$, the terms $T_{je}^k \to 0$ as $k \to \infty$. Thus,

$$\sum_{e=0}^{l} a_{le}\alpha^e\beta^{l-e} = 0 \qquad \text{for each } (\alpha, \beta).$$

For those approach vectors with $\alpha \neq 0$, we can divide through by $\alpha^l$:

$$\sum_{e=0}^{l} a_{le}\left(\frac{\beta}{\alpha}\right)^{l-e} = 0. \tag{A.1}$$

We now must consider a few cases.

CASE 1: *Assume that either $d > l$, or $d = l$ but there are $d+1$ distinct approach vectors with $\alpha \neq 0$.* In this case, there are at least $l+1$ approach vectors $(\alpha, \beta)$ with $\alpha \neq 0$. If we write $f(t) = \sum_{e=0}^{l} a_{le}t^{l-e}$, then for each such approach vector, we have $f\left(\frac{\beta}{\alpha}\right) = 0$ by (A.1). As there are at least $l+1$ such values and $\deg(f(t)) \leq l$, we must have $f(t)$ identically zero. Hence, $a_{l0} = \cdots = a_{ll} = 0$ and therefore $p_l(x, y) = 0$ for all $(x, y)$.

CASE 2: *Assume that $d = l$, and that one of the $l+1$ approach directions is* $(0, \pm 1)$. In this case, we see directly that

$$|a_{l0}| = \left|\sum_{e=0}^{l} a_{le}0^e(-1)^{l-e}\right| = 0$$

Thus, $a_{l0} = 0$ and the polynomial $f(t)$ defined as in the previous case is actually of degree at most $l-1$. Then, for the $d = l$ approach vectors $(\alpha, \beta)$ with $\alpha \neq 0$, we have $f\left(\frac{\beta}{\alpha}\right) = 0$. We conclude, as before, that $f(t)$ is identically zero. Hence, $a_{l0} = \cdots = a_{ll} = 0$ and $p_l(x, y) = 0$ for all $(x, y)$.

In all cases, $p_l(x, y) = 0$ as desired. This completes the induction step and the proof. □

# Appendix B

# Lyapunov Exponents of Upper Triangular Matrix Products

In this appendix, we prove Theorem 3.2 that when a sequence of upper-triangular $m \times m$ matrices are multiplied together, we can find vectors whose lengths grow at the same rate as the diagonal elements of the matrix product. This result can be found elsewhere in the literature, for example in [17], though it usually is presented in a more general context. This presentation of the theorem roughly follows [17] but can also be considered a distillation of a piece of Oseledec's original proof.

**Lemma B.1.** *For $k = 1, 2, \ldots$, let $A_k$ be an $m \times m$ matrix written in block structure $A_k = \begin{pmatrix} a_k & b_k \\ 0 & B_k \end{pmatrix}$, where $a_k \in \mathbb{R}$, $b_k \in \mathbb{R}^{1\times(m-1)}$, and $B_k \in \mathbb{R}^{(m-1)\times(m-1)}$. Set $S_0 = I_m$, the $m \times m$ identity matrix, and for $n = 1, 2, \ldots$ define $S_n = A_n \cdots A_1 = \begin{pmatrix} s_n & t_n \\ 0 & T_n \end{pmatrix}$, using the same block structure. Then*

$$t_n = a_n \cdots a_1 \sum_{k=1}^{n} \frac{b_k T_{k-1}}{a_k \cdots a_1}.$$

**Proof.** The proof is by induction on $n$. For $n = 1$, $t_1 = a_1 \sum_{k=1}^{1} \frac{b_1}{a_1} = b_1$. In general,

$$\begin{aligned} t_{n+1} &= a_{n+1} t_n + b_{n+1} T_n \\ &= a_{n+1} a_n \cdots a_1 \sum_{k=1}^{n} \frac{b_k T_{k-1}}{a_k \cdots a_1} + a_{n+1} \cdots a_1 \frac{b_{n+1} T_n}{a_{n+1} \cdots a_1} \\ &= a_{n+1} \cdots a_1 \sum_{k=1}^{n+1} \frac{b_k T_{k-1}}{a_k \cdots a_1}. \end{aligned}$$

□

**Definition 3.1.** *A sequence of real numbers* $\{r_k\}$ *has* **(geometric) growth rate** $\gamma$ *provided*

$$\lim_{k\to\infty} \frac{\ln |r_k|}{k} = \gamma.$$

**Theorem 3.2.** *For* $k = 1, 2 \ldots$, *let* $A_k$ *be* $m \times m$ *upper triangular matrices, and define* $S_n = A_n \cdots A_1$. *Assume the magnitudes of the entries of* $A_k$ *are bounded independent of* $k$, *and that the diagonal entries of* $S_n$ *have growth rates* $\gamma_1, \ldots, \gamma_m$ *as* $n \to \infty$. *Then there exists vectors* $v_1, \ldots, v_m$ *such that for each* $i = 1, \ldots, m$, $\|S_n v_i\|$ *has growth rate* $\gamma_i$.

**Proof.** The proof is by induction on $m$. Notice that when multiplying two upper triangular matrices, the lower $i$ rows of the product are independent of the entries of the two matrices above the lower $i$ rows. So, the induction on $m$ will start at the lower right and work upward.

Case $m = 1$ is clear; the vector $v_1 = (1)$ works. For the general case, assume that the theorem holds for all such sequences of $(m-1) \times (m-1)$ matrices. We use the partitioned notation for $A_k$ and $S_n$ from Lemma B.1. By the induction hypothesis, we will assume that by denoting by $\hat{v}_2, \ldots, \hat{v}_m$ the columns of the matrix

$$V_{m-1} = \begin{pmatrix} 1 & v_{23} & \ldots & v_{2m} \\ & 1 & \ldots & v_{3m} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix},$$

the sequence $\|T_n \hat{v}_i\|$ has growth rate $\gamma_i$ for $i = 2, \ldots, m$. (Recall from Lemma B.1 that $T_n$ is the lower right $(m-1) \times (m-1)$ submatrix of $S_n$.) In particular, for each $\epsilon > 0$ there exists a constant $K$ such that

$$\frac{1}{K} e^{(\gamma_i - \epsilon)n} \leq \|T_n \hat{v}_i\| \leq K e^{(\gamma_i + \epsilon)n} \tag{B.1}$$

for $i = 1, \ldots, m$, and furthermore, by assumption

$$\frac{1}{K} e^{(\gamma_1 - \epsilon)n} \leq s_n = a_n \cdots a_1 \leq K e^{(\gamma_1 + \epsilon)n}. \tag{B.2}$$

We will add a top row to $V_{m-1}$ to get

$$V_m = \begin{pmatrix} 1 & v_{12} & \ldots & v_{1m} \\ & 1 & \ldots & v_{2m} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$$

such that using the new columns $v_1, \ldots, v_m$ of $V_m$, the sequence $\|S_n v_i\|$ has growth rate $\gamma_i$ for $i = 1, \ldots, m$. In fact, this is already clear for $i = 1$; we just have to

make sure none of the other growth rates were changed by adding the top row of $V_m$.

Now we give the definitions of the new entries in the top row, $v_{12}, \ldots, v_{1m}$. If $\gamma_1 \leq \gamma_i$, set $v_{1i} = 0$. If $\gamma_1 \geq \gamma_i$, define

$$v_{1i} = -\sum_{k=1}^{\infty} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1}. \tag{B.3}$$

This series converges by comparison to a geometric series, because for each $\epsilon > 0$, the magnitude of the $k$th term is bounded above by a constant (independent of $k$) times $\frac{e^{(\gamma_i+\epsilon)k}}{e^{(\gamma_1-\epsilon)k}} = e^{(\gamma_i-\gamma_1+2\epsilon)k}$, using equations (B.1) and (B.2). Here we used the assumption that the magnitude of the entries of $b_k$ from $A_k$ obey a bound independent of $k$.

To finish, we need to show that for each $i = 2, \ldots, m$, the growth rate of $\|S_n v_i\|$ is $\gamma_i$. There are two cases. First, we consider $i$ such that $\gamma_1 \leq \gamma_i$.

$$S_n v_i = \begin{pmatrix} t_n \hat{v}_i \\ T_n \hat{v}_i \end{pmatrix} = \begin{pmatrix} a_n \cdots a_1 \sum_{k=1}^{n} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1} \\ T_n \hat{v}_i \end{pmatrix} \tag{B.4}$$

where we have used Lemma B.1 to rewrite $t_n$. Equations (B.1) and (B.2) imply that for each $\epsilon > 0$ there exist constants independent of $n$ such that

$$\begin{aligned} \left\| a_n \cdots a_1 \sum_{k=1}^{n} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1} \right\| &\leq \mathit{constant} \cdot e^{(\gamma_1+\epsilon)n} \sum_{k=1}^{n} \frac{e^{(\gamma_i+\epsilon)k}}{e^{(\gamma_1-\epsilon)k}} \\ &\leq \mathit{constant} \cdot e^{(\gamma_1+\epsilon)n} \frac{e^{(\gamma_i-\gamma_1+2\epsilon)(n+1)} - 1}{e^{\gamma_i-\gamma_1+2\epsilon} - 1} \\ &\leq \mathit{constant} \cdot e^{(\gamma_1+\epsilon)n} \frac{e^{(\gamma_i-\gamma_1+2\epsilon)n}\ e^{(\gamma_i-\gamma_1+2\epsilon)}}{e^{\gamma_i-\gamma_1+2\epsilon} - 1} \\ &\leq \mathit{constant} \cdot e^{(\gamma_i+3\epsilon)n} \end{aligned}$$

Since both entries of $S_n v_i$ have $\gamma_i$ as an upper bound for the lim sup of the growth factor as $n \to \infty$ (using the above inequality and equation (B.1)), and since the lower entry has $\gamma_i$ as a lower bound for the lim inf of the growth factor (again from (B.1)), the growth factor limit exists and is $\gamma_i$.

Second, we treat the case where $\gamma_1 > \gamma_i$.

$$\begin{aligned} S_n v_i &= \begin{pmatrix} s_n v_{1i} + t_n \hat{v}_i \\ T_n \hat{v}_i \end{pmatrix} \\ &= \begin{pmatrix} -a_n \cdots a_1 \sum_{k=1}^{\infty} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1} + a_n \cdots a_1 \sum_{k=1}^{n} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1} \\ T_n \hat{v}_i \end{pmatrix} \\ &= \begin{pmatrix} -a_n \cdots a_1 \sum_{k=n+1}^{\infty} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1} \\ T_n \hat{v}_i \end{pmatrix}. \end{aligned}$$

We can bound the top entry as follows. For each $\epsilon > 0$ there exist constants independent of $n$ such that

$$\begin{aligned}\left\| a_n \cdots a_1 \sum_{k=n+1}^{\infty} \frac{b_k T_{k-1} \hat{v}_i}{a_k \cdots a_1} \right\| &\leq constant \cdot e^{(\gamma_1+\epsilon)n} \sum_{k=n+1}^{\infty} \frac{e^{(\gamma_i+\epsilon)k}}{e^{(\gamma_1-\epsilon)k}} \\ &= constant \cdot e^{(\gamma_1+\epsilon)n} \frac{e^{(\gamma_i-\gamma_1+2\epsilon)(n+1)}}{1-e^{\gamma_i-\gamma_1+2\epsilon}} \\ &\leq constant \cdot e^{(\gamma_i+3\epsilon)n}.\end{aligned}$$

Just as in the first case, both entries of $S_n v_i$ have $\gamma_i$ as an upper bound for the growth factor, and the lower entry has $\gamma_i$ as a lower bound for the growth factor. Thus, the growth factor is $\gamma_i$. □

# BIBLIOGRAPHY


[1] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57:617, 1985.

[2] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, and S. Ciliberto. Liapunov exponents from time series. *Phys. Rev. A*, 34:4971, 1986.

[3] M. Sano and Y. Sawada. Measurement of the Lyapunov spectrum from a chaotic time series. *Phys. Rev. Lett.*, 55:1082, 1985.

[4] F. Takens. Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics*, number 898. Springer-Verlag, 1981.

[5] N. Packard, J. Crutchfield, J.D. Farmer, and R. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45:712, 1980.

[6] T. Sauer, J.A. Yorke, and M. Casdagli. Embedology. *J. Stat. Phys.*, 65:579, 1991.

[7] U. Parlitz. Identification of true and spurious Lyapunov exponents from time series. *Int. J. Bif. Chaos*, 2:155, 1992.

[8] R. Brown, P. Bryant, and H. Abarbanel. Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Phys. Rev. A*, 43:2792, 1991.

[9] J.J. Healey, D.S. Broomhead, K.A. Cliffe, R. Jones, and T. Mullin. The origins of chaos in a modified Van der Pol oscillator. *Physica D*, 48:322, 1991.

[10] A. Schenck zu Schweinsberg et al. Quasicontinuous control of a bronze ribbon experiment using time-delay coordinates. *Phys. Rev. E*, 55:2145, 1997.

[11] B.R. Hunt, T. Sauer, and J.A. Yorke. Prevalence: A translation-invariant 'almost every' on infinite-dimensional spaces. *Bull. Amer. Math. Soc.*, 27:217, 1992.

[12] E. Ott, T. Sauer, and J.A. Yorke. *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*. Wiley Interscience, New York, 1994.

[13] A.C. Aitken. *Determinants and Matrices, 9th edition*. Interscience Publishers, New York, 1962.

[14] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1997.

[15] B. Noble and J.W. Daniel. *Applied Linear Algebra*. Prentice-Hall, Engelwood Cliffs, 1988.

[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C, Second Edition*. Cambridge University Press, Cambridge, 1997.

[17] R. Johnson, K. Palmer, and G. Sell. Ergodic properties of linear dynamical systems. *SIAM J. Math. Anal.*, 18:1, 1987.

[18] H. Abarbanel. *Analysis of Observed Chaotic Data*. Springer-Verlag, New York, 1995.

[19] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189, 1983.

[20] M. Hénon. A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.*, 50:69, 1976.

[21] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, 1997.

[22] T. Sauer, J.A. Tempkin, and J.A. Yorke. Spurious Lyapunov exponents in attractor reconstruction. *Phys. Rev. Lett.*, 80:4341, 1998.